DOCUMENT RESUME

ED 026 971 24 HE 000 765

By-Davis, Junius A.
Applications of the Science of Measurement to Higher Education.
Duke Univ., Durham, N.C.
Spons Agency-Office of Education (DHEW), Washington, D.C. Bureau of Research.
Bureau No-BR-6-1722-32
Pub Date Apr 68
Contract-OEC-2-6-061722-1742
Note-154p.
EDRS Price MF-$0.75 HC-$7.80
Descriptors-Academic Aptitude, Academic Performance, *Admission Criteria, *College Admission, *Higher Education, Individual Differences, *Measurement Instruments, Readiness (Mental), Student Evaluation, *Testing

Part I of the report provides a historical development of admissions procedures in US colleges and universities from the seventeenth century to the beginning of the twentieth century. During this period student selection practices differed at each institution but were generally based on prescribed standards of academic readiness. The need for consistency in requirements led to establishment of the College Entrance Examination Board, which administered standardized testing across institutions to evaluate student performance (scholastic achievement) and predict grades in college (scholastic aptitude). The Educational Testing Service later became the Board's testing agent to build, administer and score examinations, report test results and conduct necessary research. Part II covers the second half of the twentieth century in which measurement emerged as a science, supplementing measures of academic aptitude and high school performance with measures of other variables such as interests, motivation, leadership, and other individual student differences. Research organizations or teams in university-based centers currently utilize measurement science to study problems such as student input factors, influential forces within college environments and their impact on students. These efforts could expand to include studies on the interaction between students and their learning environments, teaching procedures for heterogeneous student bodies, and the improvement of criteria by which students are evaluated. (WM)

BR-6-1722-32
PA-24
OE-BR

**NEW DIMENSIONS**
in Higher Education

Number 32

# APPLICATIONS OF THE SCIENCE OF MEASUREMENT TO HIGHER EDUCATION

April, 1968

The literature research review reported in this manuscript was made possible by Contract Number OEC 2-6-061722-1742 between Duke University, Durham, North Carolina and the United States Office of Education.

HE000765

POSITION OR POLICY.

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

NEW DIMENSIONS
IN HIGHER EDUCATION

Number 32

# APPLICATIONS OF THE SCIENCE OF MEASUREMENT TO HIGHER EDUCATION

Junius A. Davis

Everett H. Hopkins, Editor

U. S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

JOHN GARDNER, Secretary
Office of Education
HAROLD HOWE II, Commissioner

# ABOUT THE AUTHOR

Junius Ayers Davis is Director of the Southeastern Office of Educational Testing Service, located at Durham, North Carolina. He received the A.B. degree from the University of North Carolina, in mathematics; the A.M. degree from Teachers College, Columbia University, in guidance; and the Ph.D. degree from Columbia University in counseling psychology.

As a member of departments of psychology, education, or sociology, he has held teaching appointments at a number of institutions, including Orange County Community College (N.Y.), Princeton University, Emory University, the University of North Carolina, Rutgers University, Brooklyn College, New York University, and Columbia University. In 1957, he established, as Director of Testing and Guidance for the University System of Georgia, a centralized research and advisory service for the seventeen colleges of that system. From 1958 to 1961, he was Graduate Dean of the University of North Carolina at Greensboro, and Professor of Psychology and Education. In 1961, he joined the research staff of Educational Testing Service in Princeton, New Jersey, where, immediately prior to his present assignment, he was head of the Higher Education Research Group. He has served widely as a consultant on higher education research, with current standing appointments with the Veterans Administration, Williams College, Duke University, and the Regional Educational Laboratory for the Carolinas and Virginia.

Dr. Davis' memberships in professional organizations include the American Psychological Association, the American Personnel and Guidance Association, Sigma Xi, the Association for Institutional Research, and the American Educational Research Association. He has published widely in educational and psychological journals, with recent emphasis on admissions research, criteria of educational development, and academic climate.

## TABLE OF CONTENTS

Page

About the Author iii

Foreword v

Highlights vi

I. The Development of Selective Admissions Practices in American Colleges and Universities 1

  A Brief History of Selection Practices in U. S. Colleges 1

  Admissions in the First Half of the Twentieth Century: Evolution of the College Board 20

  The Prediction of Success in College 43

  The Colleges and Admissions Today 57

II. Applications of Measurement in Higher Education for the Second Half of the Twentieth Century: Achievement and Prospect 68

  The Emergence of a Science of Measurement 68

  The Enlarged Concern with Student Input 86

  The Enlarged Concern with the Educational Context 102

  The Measurement of Impact of Colleges Upon Their Students 114

  Directions for Future Research and Development 117

Footnotes 124

Annotated Bibliography 133

Reactions

# FOREWORD

(If and when this manuscript is published for general distribution, the Editor will gladly prepare an appropriate Foreword for the wider audience.)

# HIGHLIGHTS

Part I of this literature review provides historical perspective for the development of selection practices in American colleges and universities, as well as a review of the development and status of measurement science in its most important routine working application to higher education, namely, selective admissions into college. Part II begins with a more intensive look at the development of general measurement science, and reviews its application, particularly in the last decade, to other uses in higher education.

1. American colleges have been selective, covertly if not overtly, from the very beginnings in the seventeenth century. "Standardized" testing across colleges for selection purposes began only in the current century, in response to both secondary school and college needs.

2. Admissions testing in its present form began only in the late 1940's; the objective test for this purpose came into being for efficiency reasons, but it has persevered because of the ubiquitous relationship between tested measures of scholastic aptitude and academic performance in college.

3. A variety of attempts to supplement measures of tested academic aptitude and measures of high school performance with measures of other traits (e.g., interests and motivation) have not improved the prediction of performance in college to the extent that there the measures of these other traits are in common use.

4. Measurement as a science began only in the current century. It is marked with a preponderance of concern, during its first fifty years, with a focus on individual differences; only in the last decade have measurement researchers in higher education begun to extend the science to the measurement of social and institutional forces.

5. The last ten years marks an explosion of interest in using measurement science for study of a variety of problems in higher education. Significant factors in this explosion are the emergence of organizations with multi-college interests and responsibilities, the availability of substantial funding for massive efforts, and the use of mission-oriented teams of measurement research specialists.

6. Exciting new applications of measurement include the broader study of student input factors, of procedures for measuring important forces in the learning environments, and the prescription of elements necessary for a developing insight into the impact of colleges on students.

7. The new look for the decade ahead may well be a concern with measurement as a tool in the assessment of interaction between the individual and his learning environment, toward prescribing and validating effective teaching procedures for a variety of individuals, rather than as a tool for only sorting out those who learn quickly and readily in conventional situations and where success is measured through standard grading practices.

# I. THE DEVELOPMENT OF SELECTIVE ADMISSIONS PRACTICES IN AMERICAN COLLEGES AND UNIVERSITIES[1]

### *A Brief History of Selection Practices in U. S. Colleges*[2]

In the beginning was Harvard.

The model was Emmanuel College of Cambridge University. The year, 1638. The function of the new college was to insure a literate and enlightened clergy of native sons after those educated in England had wheezed for the last time through Old Hundredth; the institutional goal, as found in the *Statuta Collegii Harvardini* of 1642, was simply: "Considerato unusquisque ultimum finem vitae ac studiorum, cognitionem nimerum Dei et Jesu Christi, quae est vita aeterna." The course of studies, standing virtually without change throughout the first hundred years of operation, involved the learned languages and their grammars, rhetoric, and theological and philosophical disputations. Latin was not only a major subject of studies, but also the medium of communication, and students were forbidden to use their mother tongue within the limits of the college.

In this context, and against some of the issues and complexities of selective admissions today, the first known statement of admissions requirements stands as a reflection of and tribute to the simplicity and austerity that characterized New England in that period. The statutes of 1642, translated for the unenlightened by some unknown Puritan, read: "When any Scholar is able to read Tully or such like classical Latin *ex tempore*, and make and speake true Latin in verse and prose, *suo (ut aiunt) Marte*, and decline perfectly the paradigms of nounes and verbes in ye Greeke tongue, then may hee bee admitted into ye College, nor shall any claime admission before such qualifications."

Thus, the criterion for admissions involved a standard of academic readiness, in terms of area and level of scholarly achievement. The standard provided guidance for those involved in preparing students for college; it also defined the elements of a simple situational test that any college rector or tutor could administer across the maple desk toward the identification, rather effectively, of prospective students who could interact with the academic world into which they were about to plunge.

Toward the conscience vs. intellect debate today it is interesting to note that Increase Mather, president of Harvard from 1685 to 1701, failed in an attempt to have a religious test inserted in the college charter; that event and ensuing bitter controversy resulted not in a revision of admissions criteria but instead the founding of Yale by the defeated orthodox Calvinists. Yet the first laws of Yale College stated that "until they should provide further, the Rectors and Tutors should make use of the orders and institutions of Harvard College." Then, as now, the changing of educational practices was about as difficult to achieve as the moving of a graveyard; the underlying vested interests, though with some small differences in packaging, are in both instances deeply entrenched in the common dust of origin.

Therefore such blissfully defensible admissions criteria as those cited were maintained by and large through the eighteenth century by Harvard, and emulated with only minor variations by the twenty-two other institutions of higher education that had appeared by 1800. Although these new institutions were a product of sectarian differences and regional concerns, their reliance on mother Harvard for faculty as well as a model of necessary statutes, and on the grammar schools that existed solely as college preparatory institutions, prevented drastic revision from appearing necessary. Other subjects such as arithmetic and the sciences had not yet made substantial entry into the curriculum of either college or grammar school. The communality in origin, purpose, and curriculum (in which no flexibility could be either afforded or tolerated) preserved the uniformity in the stated requirement (although some variation among institutions in the required Latin authors had crept in by 1800, and some institutions outside the Boston area had added a requirement for the rules of vulgar arithmetic). Still another factor which permitted the standard requirement to stand without much modification was that there was "a general laxity of enforcement of the stipulated regulations for admission, and the [oral] examination was apparently a flexible and informal affair."[3]

For America, the last half of the eighteenth century was a period of sweeping and dramatic change. The frontier, principally notable before for bears and buffalo, began to acquire barrooms and bawdy houses, and the way was paved for the exploitation of the rich natural resources. Political independence came about in a climate wherein those most successful as architects for the breakdown of class and caste, or for equality of access, won the responsibility of managing the country. The state, with something equally commendable but considerably more immediate to offer than the Church, experienced a separation from the Church, and the Great Awakening was on.

But colleges then--as are colleges sometimes now--seemed loathe to change. Their scholars and tutors went on chattering in Latin and Greek, and the curriculum remained virtually unchanged until other circumstances to be examined later broke through with the Civil War. The real impact of the new order that would ultimately play the major role in the development of admissions requirements was on new forms of pre-college or non-college education.

First came the academies, probably starting with Philadelphia Academy in 1753. Beyond Latin and Greek they offered such subjects as English grammar, geography, algebra, geometry, natural philosophy, astronomy, music, composition, oratory, bookkeeping, logic, and virtue. The academies saw no reason apparently why the minister-tutor or the grammar school should be the sole springboard for entry into college; some of their students not only desired to continue their education at " higher" levels, but also sometimes excelled those who entered by other routes. The colleges responded by expanding and intensifying the specification of the traditional subject masteries for admission, and by gradually noting (and requiring) achievement in some of the new subjects. By 1807, for example, Harvard's requirements read:

> No one shall be admitted, unless he be thoroughly acquainted with the Grammar of the Greek and Latin languages, in the various parts thereof, including Prosody--can properly construe and parse Greek and Latin authors--be well instructed in the following rules of Arithmetic, namely, Notation, simple and compound, Addition, Subtraction, Multiplication, and Division, together with Reduction and the single Rule of Three; have well studied a Compendium of Geography, can translate English into Latin correctly--and have a good moral character. Each candidate shall be examined in the Grammar of the Greek and Latin languages, and in any parts of the following Greek and Latin Books, with every part of which he must be acquainted, namely, Dalzel's Collectanea Graeca Minora, The Greek Testament, Virgil, Sallust, and Cicero's Select Orations.

The second impact of the new movement was the emergence in the nineteenth century of the public high school. Beginning in 1821 as the "people's college," these institutions were not initially designed to provide college preparatory work, but rather practical terminal courses to ensure a literate population. But as tax-supported education and insistence on state responsibility in ensuring equal access to education spread, college preparatory subjects were added. Bowles,[4] in his review of "the evolution of admissions requirements," attributes great significance to the Kalamazoo case of 1874, wherein "it was held . . . that the state could act within its rights 'to furnish a liberal education to the youth of the state in schools brought within the reach of all classes.'" The forty high schools in 1860 grew to 2,500 by 1890.

Broome's review, though restricted to the major colleges, gives the reaction of the institutions of higher education over the period from 1800 to 1870. Not only were there invasions of new subjects in the preparatory experience, but also the intrusion of new teachers--many of whom, in the rapid expansion of lower education, did well to keep a jump ahead of their students. No longer content that applicants "make and speake true Latin," Harvard examined, in 1869, candidates for admission to the freshman class "in the whole of Virgil; the whole of Caesar's Commentaries; the Orations of Cicero, included in Folsom's, Johnson's or Stuart's edition; Latin Grammar, including Prosody; and in writing Latin." Some hundred miles away, Yale examined students the same year in "Latin Grammar, including Prosody; Sallust--Jugurthine War, or four books of Caesar; Cicero--Seven Orations; Virgil--the Bucolics, Georgics, and first six books of the Aeneid; and Arnold's Latin Prose Composition, to the Passive voice (first XII chapters)." The requirements in Greek grew even more strenuous and varied. Yet the college curricula had changed little up to this time. The admissions procedures most exactly reflected the harsh look downward into preparation rather than the hopeful look upward into promise; admissions requirements had become not so much a tool for the guidance of pre-college work as a weapon to impose the college perception of what indeed good preparation had to be. The extensiveness of the new requirements may also have reflected a tactic for resisting, in the name of quality, the diverse curricular innovations outside the traditional college preparatory courses.

The War Between the States, the technological revolution, and the increasing clamor for public education, all had a profound effect on the 200-year-old pattern of college curricula. In the middle of the nineteenth century, the state of Massachusetts had withdrawn its financial support of Harvard when a committee of the state legislature found that an outdated curriculum failed to meet popular needs.[5] Enrollment problems attributable to the war did not ease immediately following the war's end. The prospective students wanted something other than the theoretical and philosophical excursions backward into the classical world; they were anxious to learn how to exploit new ideas and techniques in the present and emerging world. Though new colleges were being developed in great numbers by a variety of sects, it was the sect, not the mass of applicants, that desired an educated clergy, and after a brief period most of these institutions collapsed because of financial problems and want of students.

As early as 1830 there had been some experimentation with new college curricula. In that year Columbia initiated the " Scientific and Literary Course" which, according to the statutory enactment, was established with a " view of rendering the benefits of education more generally accessible to the community." For admission to that three-year non-degree program, students were required to have a grammatical knowledge of French, as well as meeting the usual requirements in mathematics and geography. In the 1850's Brown, Harvard, Yale, Dartmouth, Rochester, and Michigan introduced new non-classical degree programs. However, as these came because of pressures from outside the colleges rather than from pressures within, both curricula and admissions requirements (the latter in the pattern borrowed from the classics and now synonymous with quality or academic or institutional respectability) developed with an infinite variety of specific preparatory prescription. This was the college's way of attempting to make such new programs reputable in its own eyes, or worthy of its attention and grudging blessing, with a new degree (Ph.B. at Brown and Yale, and B.S. at Harvard, Dartmouth, Rochester, and Michigan) as the face-saving rationalization to the higher academic community.

But both the new programs and their admissions requirements need to be viewed in terms of a major principle stated by Bowles:[6] "The enforcement, or qualitative aspect, of entrance requirements is determined by higher education in response to applicant supply and demand and with little or no reference to the attitudes and objectives of secondary education." The new college programs were not so much founded on deep convictions as to the needs of a new society, but on urgent pressures to capture enough students to permit the institutions themselves to survive. This observation is of tremendous importance to those who would keep a sane head and a perceptive eye on the modern struggles between preparatory and college forces; it is also fundamental to those who would understand that the admissions problem may represent at first blush the tool for repairing difficulties, but that it then becomes a dilemma with one horn piercing the maintenance of collegiate or qualitative standards and the other horn firmly planted in the necessity to adjust levels to permit sufficient enrollment.

Thus, the explosion of public education, the enrollment and survival crises in the colleges, and the technological revolution sounded the death knell of the Colonial College, with the War Between the States and the ten years thereafter affording the period of break and transition. Old institutions that would continue, and new colleges that would assume a firm foothold, were forced to be responsive to the non-classics, the mushrooming sciences and technologies, supporting mathematics, the modern foreign languages, and the other subjects introduced by the secondary schools.

The emergence of the modern American college in 1870 carried with it the colonial admissions patterns transmuted to the new subjects. With college study today rapidly becoming the prerogative of all, any modern scholar of more existential than historical bent who believes current controversies and problems are at a peak never before experienced needs only to listen to the hue and cry of the period from 1870 to 1900 in America. In his account (which reflects the biases of the private preparatory schools or academies), Fuess has stated:

> For the preparatory schools the uncertainty was both ludicrous and tragic. As [Nicholas Murray] Butler said, "If Cicero was prescribed, it meant in one place four orations and another six, and not always the same four or the same six." When some colleges demanded Greek Composition or Latin Composition, a school's classical department had to form special sections to meet the need. Each college, furthermore, held its entrance examinations to suit its convenience, with the result

> that the time schedule of a school like St. Paul's or Newton High School during the spring term was disrupted. One such group of examinations was set on the day of a school's most important baseball game, and the local protests were violent. Dr. Cecil F. P. Bancroft, principal of Phillips Academy, Andover, complained pathetically in 1885 that "out of every forty boys preparing for college next year we have more than twenty Senior classes." . . . The written [entrance] examinations themselves, often dictated hastily by professors with small knowledge of student psychology, were unscientific and varied in difficulty from year to year and from college to college.[7]

For the public schools, the situation in readying students for college must have been even more ludicrous. The high schools, showing the usual tendency of new public institutions to try to be all things to all people, did take on the task of preparing students for the traditional programs of classical studies as well as for the newer, more practical areas of college work or for entry directly into work. The older, more traditional areas probably attracted both the better students and the better teachers; the newer, more pragmatic areas fared less well on both counts. These newer areas also sustained more suspicion from the colleges and entertained greater efforts, through admissions, to control their quality. The high schools grew in importance, however, and this new market produced the observation of Broome at the start of the twentieth century that "the history of college admission requirements for a quarter of a century has been a series of concessions to the high schools."

These concessions were not easily granted; indeed, to give too much emphasis to them is to underrate grossly the impact of the college, principally through its admissions standards, in providing the secondary schools with a mark at which to shoot. Even the nature of the tests themselves may have had some desirable impact, for as the range of facts to be sampled on admissions examinations increased, the colleges necessarily began to focus on more general evidences of learning. Broome noted " the emphasis placed on sight translation in the language examinations, the growing importance of English composition, of the solution of original problems in geometry, and of independent experimental work in science." As a result of these tendencies in college admission examinations, he concluded: " There has been a significant revolution in preparatory school methods of teaching, a shifting of the emphasis from stultifying memoriter work to that more quickening sort which calls for independent thought and constructive ability."

There were, however, other solutions to the problem of how to transmit the mold of earlier testing in the classical requirements into the broader arena of the expanding body of subject matter. One procedure was the adoption of a variety of alternatives in the subjects on which students might be examined. A bolder plan was initiated by the University of Michigan in 1870. The Calendar for that year stated:

> Whenever the Faculty shall be satisfied that the preparatory course in any school is conducted by a sufficient number of competent instructors, and has been brought up fully to the foregoing requirements, the diploma of such school, certifying that the holder has completed the preparatory course and sustained the examination in the same, shall entitle the candidate to be admitted to the university without further examination.

One can well imagine the furrowing of shaggy brows in Cambridge, New Haven, and Princeton, particularly as Indiana University, the University of Wisconsin, and the University of California followed suit in the next fifteen years (the Eastern colleges had already begun the "certificate system," whereby those particular principals whose wisdom and rigor were certain from the fact of graduation from the mother institutions were sometimes allowed to vouch for their candidates). The Michigan or "diploma system" was recognized as superior to the Ivy League "certificate system" even by President Eliot of Harvard, who saw the attendant procedures of the former--involving inspection of the secondary school by the college faculty--as a means to greater communication, interaction, and stimulation. Thus, the diploma system assumed the previously noted function of admissions requirements as guidelines for preparatory work, yet it may have achieved this more by friendly cooperation than by the superimposed threat of the test or examination standard. Whatever stand--or strategy--any modern critic would be inclined to take, Michigan, Indiana, Wisconsin, and California seem one hundred years later to be viable institutions of higher learning.

The last quarter of the nineteenth century was to see an important characteristic of American educational systems develop--predominantly as a function of the diversity of admissions requirements and solutions. This was the formal organization of professional associations representing both colleges and secondary schools. The colleges, isolated from one another, harboring delusions of self-sufficiency, or priding themselves on their own particular brands of wisdom, were not in any mood to ease the admissions preparation beyond relaxing standards when necessary to keep classes open (it is a curious fact that no instance is known where professors closed their books and went home for failure of available students to meet a predetermined standard of excellence). This did not help the problems the secondary schools were facing. The first school-college organization was the New England Association of Colleges and Preparatory Schools, established in Boston in 1885 with the aim of "the advancement of the cause of liberal education by the promotion of interests common to college and preparatory schools."[8] What these interests were became apparent from the outgrowth from this organization in 1886 of the Commission of Colleges in New England on Entrance Examinations, representing all but five of the colleges in New England, with its more precise aim "to devise means for securing greater uniformity in college admissions examinations."[9] By 1897 there were twenty-three colleges and other educational associations devoting time and attention to the problem of securing a workable uniformity among colleges.

These different organizations had varying degrees of success within their regions of influence, but they met with enough success to demonstrate that uniformity within regions was not sufficient and to attract national efforts. The National Education Association, an organization more representative of public school interests than any other, appointed in 1892 the "Committee of Ten" to look at the problem on a national scale. The recommendations of this group were studied by an appointed "Committee on College Entrance Requirements," which involved nearly 150 experts in both secondary and higher education working together for more than four years. Their final report, in 1899, is a masterpiece of educational architecture. That is, it did not attempt merely to impose a prescription for entrance requirements, but also attempted to strengthen through guidelines for the secondary schools their preparatory efforts (e.g., "we recommend an increase in the school day in secondary schools, to permit a larger amount of study in school under school supervision"). The uniformity was not to be gained by a common prescription, but through a system of units among various common subjects, with colleges to name the most crucial options within these units.

The work of the NEA and the regional groups is a testimony to the seriousness of the problem, and their recommendations, in terms of what was known at that time, were both sincere and sound. But what fiery guardian of sacred standards on any college campus has been known to demur to a committee representing those very agents he is dedicated through his standards to snare by the ear and lead to higher things? It remained for the recommendations to be implemented.

At the very first meeting of the New England Association of Colleges and Preparatory Schools in 1895, President Eliot of Harvard had suggested the notion of a common examining board, an idea that had fallen on deaf ears among his Harvard faculty in 1877. Professor Butler, later to become President of Columbia College, introduced in that faculty in 1893 such a resolution, which was passed by a unanimous vote. But other institutions did not rush to examine and copy the Columbia requirements, and it remained for Dr. Butler to transport the notion to a meeting of the Association of Colleges and Preparatory Schools of the Middle States and Maryland in 1899. That association resolved to urge the early establishment of a joint college admission examination board, to exact agreement among the member colleges as to each subject required by two or more colleges, to hold uniform examinations in June of each year, to empower the board to name secondary school representatives to serve with it, and to request the member colleges to accept the certificates issued for satisfactory performance on the tests in lieu of the institutions' own examinations. Dr. Butler gave not only of his counsel, but also space for the offices what was known at that time, were both sincere and sound. But what fiery guardian of sacred standards on any college campus has been known to demur to a committee representing those very agents he is dedicated through his standards to snare by the ear and lead to higher things? It remained for the recommendations to be implemented.

At the very first meeting of the New England Association of Colleges and Preparatory Schools in 1895, President Eliot of Harvard had suggested the notion of a common examining board, an idea that had fallen on deaf ears among his Harvard faculty in 1877. Professor Butler, later to become President of Columbia College, introduced in that faculty in 1893 such a resolution, which was passed by a unanimous vote. But other institutions did not rush to examine and copy the Columbia requirements, and it remained for Dr. Butler to transport the notion to a meeting of the Association of Colleges and Preparatory Schools of the Middle States and Maryland in 1899. That association resolved to urge the early establishment of a joint college admission examination board, to exact agreement among the member colleges as to each subject required by two or more colleges, to hold uniform examinations in June of each year, to empower the board to name secondary school representatives to serve with it, and to request the member colleges to accept the certificates issued for satisfactory performance on the tests in lieu of the institutions' own examinations. Dr. Butler gave not only of his counsel, but also space for the offices of the Board, and on November 17, 1900, the College Entrance Examination Board was born.

In concluding his historical discussion three years later (through argument that should be examined in the original), Broome showed that recognition of dangers resident in the new system permitted a way around them from the very beginning. For example, an outside group taking over a responsibility of the individual college had to do a creditable job both in constructing examinations and in evaluating performance if the new system were to survive. An examination writer in a college, in the press of autumn business, could use hastily dictated exams and browse briefly through them, but now he must prove his mettle and academic integrity. The secondary school representative at the Board, sitting with his college counterpart, could neither show weakness to that counterpart nor mercy to the candidate. Uniform statement of the definition of admissions subjects was assured, together with a means of enforcing these definitions. And finally, it was recognized that each college could preserve its own brand of integrity by doing what it wished with the quantitative results: the Board would report level of performance, define (of course) a "passing" level, but leave the college free to demand higher or accept lower levels. That these assumptions proved viable was demonstrated in 1966 when the College Board, now with over 630 member colleges, expended more than $18 million to provide testing programs, a variety of admissions-related services, conferences and publications, and a varied program of research and development (with almost $1 million invested in the latter category).[8]

It would be well at this point to summarize the functions and roles that admissions requirements and procedures served in their evolution to 1900, the problems they solved and the problems they raised, and the forces controlling their establishment. Prescriptions for selective admissions started as a gentle set of guidelines for preparatory agents toward the specification of essential levels and the nature of prior learning needed for orderly progression in the new learning environment. With the appearance not only of formal pre-college institutions to conduct preparation but also of a variety of these institutions, admissions requirements moved from providing guidelines to influencing, indeed controlling, the content, quantity, and quality of pre-college studies. That this responsibility could not effectively remain with the colleges has been seen as a product of upper academic conservatism and resistance to change, particularly that from the addition of new disciplines or content areas, of economic and social pressures on the colleges for survival, of the growth of knowledge and technological change, of the weight of the public school population, and of the social and economic utility of the emerging subject matter involved in the people's colleges or public high schools. The administrator of the blow for quality received a heavier one in return. The preparatory schools, particularly the public schools, emerged in the driver's seat, and the most effective resolution of the matter of college-influenced qualitative standards came not from the rigid imposition of controls through screening procedures but from efforts of college people working with preparatory people toward operational communalities.

In the struggles to 1900, one may also discern the fact that admissions requirements were frequently perceived as a means for the college to maintain status or its own brand of reputability. Yet one man's status symbol is another man's poison, and not all those institutions or vested interests among the faculties who desired status actually earned it. The pure weight of maintaining, year after year, individually constructed examinations led frequently to inferior or faulty samplings of subject matter content; and some of the areas through which status could be expressed appealed only to the cloistered proponents rather than to the public on whom they depended for students. Also those with a particular item to sell sometimes found the market had changed.

Finally, it appeared that for administrative convenience as well as for orderly control by lower education of the preparatory experience, the responsibility for designating and administering the academic content of the admissions requirements must be passed on, if any were to survive, to forces outside the individual colleges. These emerged in two basic forms: the recognition and use of the evaluation of the student by the secondary schools, and the establishment of a new institution to represent, for a collection of colleges and preparatory schools, their common needs and interests. This institution, appearing as the College Entrance Examination Board, showed signs of succeeding by efforts (1) to represent the two partners, (2) to rely on the best scholars from the ranks of the partners to determine substantive content, and (3) to refrain from a common prescription of what content and levels each college could or could not tolerate.

## Admissions in the First Half of the Twentieth Century: Evolution of the College Board

The development of our selection practices in the first half of the twentieth century is probably best traced through the development of the College Entrance Examination Board, with attention ultimately being given to those institutions it served and those it did not (and, in the latter case, what these institutions did). Probably no single organizational or consciously contrived administrative force has had a more sweeping impact on higher education than has this organization.

Several factors in the initiation of the Board have already been alluded to, but should be recounted. First, the Board grew out of needs for enough consistency in college admissions requirements that the secondary schools might have a reasonable chance of preparing any able student for a range of institutions of higher education. The chief architects were not only the presidents of two of the most reputable and visible institutions of higher education, but also educational leaders of great substance who were recognized then as they are today. Their plan called for involvement of the most respectable academicians from both higher and lower education; the operational focus was to help define, rather than impose or enforce, academic standards, and to provide careful, sound, and fair evaluations of student performance. The Board was to be a member organization, with policy, control, and activities to be determined by representatives of participating colleges and secondary schools. And last, but not least, it would assume the operational burden of testing candidates at locations across the country, and relieve the colleges of construction, administration, and evaluation pressures.

The early history and development of the Board has been traced in an intimate, folksy account by Claude Fuess.[9] Although many errors of fact crept into this stream of personal reminiscences, and there are many important omissions, the flavor of the Board's early operation in donated space at Columbia University comes through vividly. Nine subject matter areas were agreed upon (chemistry, English, French, German, Greek, history, Latin, mathematics, and physics), and forty tests covering various courses within these areas were prepared for administration to students at the close of the year in which they took the particular courses. Guidelines and standards from the appropriate professional associations representing the discipline were consulted where such were available. A committee of examiners was appointed and a "Committee of Revision" (the nine examiners with the five representatives of the secondary schools on the Board) was established to review together the first examinations produced. In June, 1901, 973 candidates were tested in sixty-seven centers in the United States and two in Europe. The 7,889 papers written were forwarded to the Board, where thirty-nine carefully selected readers sat around tables in the Columbia University Library to evaluate them. Both the anticipation of this chore and the fact of physical meeting required formal consideration of evaluative guidelines and procedures; the task of reading, the drudgery, the points of debate, and the humorous answers sometimes encountered gave to the Board what was to be its distinctive personality and flavor for the next forty years. This was the in-group of scholars, faced with a reasonable task to give them focus, and with the responsibility of defining the true substance of intellectual development in their disciplines.

Not all institutions rushed to join the new Board. Only Columbia, Barnard, and New York University abandoned their own examinations that first year. Eliot's faculty at Harvard voted with no dissent that it was " inexpedient" to rely on the Board's certificates, and Yale trusted only its own faculty to evaluate the papers. But the Board's evaluation of the papers was, if anything, obviously too severe to permit the criticism of laxity (40.7 percent of the papers were judged in the failure category), and the examinations themselves were quite obviously superior, with the time and thought given them, than most of those produced by the individual colleges.

In 1903 the Board's constitution was amended to include three new school-college associations (New England, North Central, and Southern States) to participate with the Middle States. Harvard joined as a member institution in 1904, together with Western Reserve; Williams and Smith joined in 1907, Dartmouth and Wesleyan in 1908, Yale in 1909, and Princeton and Amherst in 1910. Joining meant sending an official representative to the annual meeting, and was not tantamount to accepting the examination program. It was not until 1915 that Harvard, Yale, and Princeton agreed to use the Board tests as substitutes for their own. Even then, these institutions wanted the Board to draw up their own tests for September (rather than June) and found that they had to furnish their own readers at that busy time.

Though Fuess' account does not say so, it would seem a good guess that other factors were at work to make the experiment successful. The early subject matter experts were key scholars from key institutions; students electing the examinations came from the pace-setting preparatory schools (whose teachers had been involved), and colleges began to take a second look when the best representatives of their faculties returned to join in the local policy debates. By 1910 the examinations were taken by 3,731 students, with 1,626 from New England schools, or 1,968 who desired to attend New England colleges.

There were, of course, mistakes made. The first examinations were not of even quality, and some that did not prove out too well gave particular pause to the readers. A policy of announcing names of top-scoring students backfired (who, indeed, could make a 100 in history, asked teachers at the schools with no such candidate). The professional associations that had been counted upon for substantive advice as well as status proved in general to have no members really interested in secondary school or freshman-level certification, and consequently were of little service. There also proved to be no effective way to resolve the question of absolute but equivalent levels of achievement from subject to subject against the varying standards of the representatives of the separate disciplines. (In 1914, for example, only 32 percent of all the candidates in American history received a passing mark; the Board review committee concluded that the reason was inadequate preparation, but had this writer been a secondary school history teacher at that time, the next 3,000 words would be devoted to demolishing that explanation.) And, for the first twenty-five years, the Board operated on deficit financing.

As interest grew, however, the Board found itself confronted with demands for examinations in additional areas: in fact, in 1902, examinations were added in Spanish, botany, geography, and drawing. By 1916 there were clearly too many bits and pieces to be manageable, and the Board resolved the issue by introducing "The New Plan," consisting of four comprehensive examinations all to be taken in the senior year. These were the tests requested by and tailored for Harvard, and agreed upon by Princeton and Yale; they were to be accompanied by the report of the student's high school average.

Two aspects of the New Plan are important. First is the change in form of the tests. The general or "comprehensive" nature eased the problem of specific content (in reducing the number of tests required to cover a subject area to one), making test-making and administration more manageable. Also, and more important, the move involved a recognition that tests going beyond memory of specific factual content into the understanding of basic relationships or the ability to perceive new relationships might be more defensible. These were perceived, in that day of Binet, as a step toward power or mental ability tests, although they were more closely akin to today's achievement tests than to scholastic aptitude as we know it now.

The second important aspect of the New Plan was the provision, through recognition of the high school average, for taking into account the judgment of the secondary school teachers personally acquainted with the student and his work. This, too, was a move away from the absolute faith in only that standard which the college representative might prescribe.

Although the Board continued to offer the old tests and services along with those of the New Plan for a number of years, it was the New Plan that foreshadowed the shape of things to come. The reasons may have been the pure weight and unmanageability of the old examinations; indeed, the Board had found in the burgeoning numbers of different examinations and candidates a reflection, rather than an easing, of the multiplicities that had led in large part to its formation. Another reason may have been the underlying truth that any teacher, at any level, must have a personal say and concern for how his students are to be evaluated for the work under his direction; or, to say it simply, that the teacher cannot afford for others to do all his thinking and planning for him (shades of the modern criticisms about teaching machines!). But the most likely reason of all is probably given in Chauncey's[10] comment about this period some years later:

> But as time passed, the colleges discovered that detailed mastery of a large number of individual subjects was

> not as important as it had been thought to be and that the school record was at least as good an index of success in college as the examination record. [Emphasis added.]

In other words, the colleges were discovering that what they wanted were students who could perform well academically, and that the report of the previous teacher was as good or better for this purpose than the test. The age of prediction of performance was about to begin.

Even with the New Plan, there was a long way to go before approximating modern experience. Chauncey also stated in his 1947 review:

> These [New Plan] comprehensives proved to be good predictors of college success if--and a significant if--the student attended a school which organized its work by the College Board's published syllabus and which gave constant drilling in essay writing on the prescribed subjects.

In other words, the goal of using tests to determine ability to handle new material rather than simply reflect old specific acquisitions had not yet been reached.

This matter caused few problems at first, for before 1920 the Board had concerned itself with the prestige colleges and the expensive private schools from which these colleges drew the bulk of their students. The 1920's, however, were the years of the postwar boom, which meant a consequent deluge for higher education. The deluge involved a mass of applicants from public high schools. These schools and their students scarcely suspected the existence of the Board, much less were they familiar with its syllabi. Also, although the private schools were malleable in their dependence for existence on the benevolence of the colleges, the public schools were in no position or disposition to look to Harvard as the only truly divine source of wisdom about curricula.

As noted earlier, colleges in America of whatever quality have not been prone to wither away because of lack of high-level students when there is a large market of potential applicants at lower levels. Neither has it been easy to lower standards, and the colleges that were to remain with the best of both possible worlds needed a way to identify and attract that segment of the large market that would be most likely to survive or do credit to the college.

In 1924 Professor Brigham of Princeton University was appointed as the chairman of a Board commission charged with developing a new test of scholastic aptitude. This was, of course, a time when those psychologists who had worked during World War I with the new mass tests, the Army Alpha and Beta, were back on their campuses, flushed with enthusiasm and seeking new fields to conquer. Brigham did produce a scholastic aptitude test, cast in "objective" format. In 1926, the Board agreed to experiment with it, and administered it without charge, sending the information to the colleges for guidance or research only.

It was a new educational consideration that provided the first entry for Brigham's new test. Harvard, Yale, Columbia, and Princeton decided that it would be desirable to attract students from other parts of the country, and so established regional scholarship programs. The scholarship applicants, coming mainly from public high schools, did poorly on the regular examinations of the Board, and there was the additional difficulty that the Board examinations, with results not available until July, came too late.

Henry Chauncey, who was responsible for the scholarship program at Harvard, was attracted both by the first experiment with the new objective tests in the Carnegie study in Pennsylvania by Learned and Wood[11] and by Brigham's work. The new objective tests could be administered quickly (the conventional tests of the Board then required a week of writing) and handled efficiently; they promised, in the absence of reliance on specific subject matter, a fair base regardless of the secondary school program. Thus, in 1937, Harvard, together with Yale, Princeton, and Columbia, asked the Board to prepare a special series of objective examinations to be used in the selection of scholarship students. This first scholarship series consisted of a Scholastic Aptitude Test (with verbal and mathematical components) and a battery of achievement tests of which the applicant was to take three.

In those days, however, award of scholarship and award of admission were two different administrative actions; the latter still depended on the regular entrance examinations. Brigham's work had tested, of course, the ability of his tests to indicate level of future academic performance; but the scholarship officers joined vigorously in this enterprise when some of their prize scholarship winners were turned down a few months later in the admissions office. The seemingly esoteric statistical work of Brigham was now studied carefully; some experimentation was agreed on in the colleges to test the new type of index of promise against the old type of index of readiness. And, against the criterion of level of academic performance in college, no advantage of the old-type tests could be found.

How long it might have taken for the Board and colleges generally to take such evidence into account on its own weight is not known, for it was another train of events that led to the program as we know it today, a train which started moving at 1:07 p.m. Eastern Standard Time on December 7, 1941. The colleges responded immediately to the war pressures with round-the-year programs and new classes to be admitted in June, 1942. In the resulting clamor, the old-style essay tests had to be abandoned (it was assumed for the duration) in favor of the more efficient model of Brigham. Second, the Board, with a psychometric laboratory now at Dr. Brigham's Princeton University, felt compelled to contribute all its talents and facilities to the war effort, and thus made the laboratory available for government needs for tests. The annual reports of those years showed, both in dollar income and in numbers of candidates tested for various college or non-college war training programs, that this work boomed. But not only was capability and momentum acquired in the laboratory: the technicians acquiring the heavy testing experience insisted on researching the effectiveness of the tests, and both improved their product and acquired substantial proof of results.

The essay tests, except for recurring attempts to contrive English Composition tests that could be graded reliably (a quality the new young breed at the laboratory insisted upon), were never to return. The patter of summer excursions of readers to Columbia had been broken, and the numbers now involved were too heavy to permit the former style of operation. Not only had the test technician matched the old subject matter expert at his game and against his ultimate performance standards, but he had also contrived a more efficient system.

Other advantages were noted. In the Report of the Executive

Secretary of the Board in 1945, John Stalnaker observed:

> In 1900, the problem was to keep out all who had not undergone very specific preparatory training. Such training was necessarily restricted to those able to attend the few secondary institutions which devoted themselves to the whims of distinguished universities. Today the problem is to attract the intelligent, apt pupil regardless of where or how he got his training and almost irrespective of what his school has been offering on its curricular menu.[12]

Following this vein, Stalnaker then noted that the Brigham brand of test was an "accurate index of pupil ability, rather than a means of controlling the curriculum." Again, the bases for Bowles' 1956 observation that the secondary schools and candidate supply were the most powerful factors in admissions procedures can be detected. But the fact remained that the new tests gave the colleges what they wanted--good students. Chauncey discussed the new look in a 1947 speech delivered before a midwestern group of college admissions counselors:

> But we have diverged not at all from our original goal of assisting admissions officers to do their job and of conserving and enriching the human resources of this country by helping to ensure that the best students go to college.... The fixed star which guides our present course I might call the star of freedom. Or I should say freedoms, for there are three. Freedom from bias in favor of any group of students, freedom of subject matter and teaching methods in the schools, and for the colleges, freedom to use the scores on our tests as they see fit.[13]

To maintain an examination program of this sort required a different kind of staff from that employed by the Board before 1941; also demanded was a substantial ongoing research operation. The laboratory in Princeton had the nucleus of staff and style (in 1946 the Princeton laboratory had more than one hundred employees, while the Board's office at Columbia had become a virtual mail drop). It remained for a Harvard president, Dr. James B. Conant, to discover once more the bold and creative solution and drive it home. As chairman of the Carnegie Foundation's Committee on Testing in 1946, he proposed that the testing functions of the Board, the American Council on Education, and the Carnegie Foundation for the Advancement of Teaching be consolidated and placed with a completely new organization, to be built from the resources of the Board's Princeton laboratory. In 1947 the charter for the new organization, to be known as Educational Testing Service, was granted by the State of New York; funds, equipment, the continuing non-college or government contracts, and staff were provided or already held to assure a splendid start. Thus, Educational Testing Service, the " testing industry" as critics then and now are likely to call it, came into being. For the Board, ETS would be used as its testing agent (to build, administer, and score tests, and to report test results) and would be paid for services rendered. It would also conduct necessary research.

Clearly, the Board could now have faded into the underbrush on Morningside Heights, leaving the brave new breed of psychologists and statisticians to rule the entrance testing business. But appointments of great significance happened. Dr. Frank Bowles became

Director of the Board, and Dr. William C. Fels its secretary. Both Bowles and Fels were men of too great ability and vision to let the passing of an era and a function displace any consideration of new contributions. With the new ETS free to handle the mechanics as well as the theoretical problems of measurement, the Board could now turn its attention to admissions problems and philosophy beyond that of affecting the curriculum of the secondary school or feeding only the vested interests of professors in the various disciplines. These new areas of service are still very much in process of formulation (some are dealt with in detail later in this review), but a few examples here will suffice: the initiation of a program for determining, by objective and confidential means, the financial capability of parents, with the aim of providing guidelines for scholarship aids (1954); the initiation of a publications program for disseminating information useful to college admissions officers and pre-college counselors; and the establishment of a grants program for support of general research (to be conducted by agencies or individuals across the country), thus expanding skills or points of view represented by ETS, its prime contractor. In short, the Board turned its attention to societal and managerial problems of admissions, relegating the testing science to the specialists.

The foregoing description of the development of the College Board omits a large part of the picture of selection for higher education in the first half of the twentieth century. This is because the Board had traditionally concerned itself with a distinct class of institution, representing only a relatively small proportion of the institutions of higher education (in 1950, less than 10 percent of the member colleges were tax-supported colleges or universities). True, the Board clientele were the influential, pace-setting institutions, important as models in every aspect of functioning; true also that these were the institutions concerned most directly with selection as a discrete administrative act (that could afford to be selective), i.e., as a function of the numbers of applicants vs. the source and amount of operating funds that probably determine optimal size. But the great mass of tax-supported institutions, or those newer private institutions that could not afford much selectivity during this period, need also to be examined: for these institutions, developing in many instances as a product of distinctively American trends and needs, have accounted for the large bulk of our college graduates, or our pools of top-level manpower.

The public colleges and universities, and many late-coming private institutions hungry for students, have generally been described as "open-door" institutions. This is, of course, not entirely true, because admissions then and now have been controlled in at least three ways: (1) by some operational concepts that define for whom the institutions are appropriate (e.g., for high school graduates, for sons and daughters of taxpayers in a given state, etc.); (2) by other more subtle forces that attract some kinds of students and dissuade others; and (3) by the amount of funds invested in higher education within a region (e.g., supply and demand considerations). In short, these institutions were not truly "open-door" any more than the concept of "education for all" would mean that college-level studies are appropriate for all the population. Instead, these were the institutions that generally relied on public-defined standards _at the point of admissions_ rather than on admissions examinations defined and controlled by the college.

The most popular and universal "public" standard was completion of secondary school studies. As a further safeguard, some kind of accrediting standards were also applied to the secondary school, as a form of guidance and control of that class of institution (these were set, typically, by state public-school administrative control agents or by a private regional accrediting association). Since 1865 New York State had monitored and certified quality and level of secondary school preparation through their Regents examination system, thus focusing on the student directly and the system indirectly. In most instances, the characteristic feature of the accrediting _standard_ was its establishment and control more by secondary school than by college interests.

How, then, were the individual colleges free to determine their own goals and standards? That such a system did not produce a mass of homogeneous institutions, or institutions that varied directly as a function of the quality of secondary schools in the drawing area, is evident from the great diversity of types and levels of institutions that developed during the first half of the century.

Some institutions were blessed with a combination of limited support, heavy pools of potential applicants, and an administration sensitive to societal needs as well as to internal college demands from competent faculty. The best example of such an institution is probably the City University of New York, whose students, at the point of admission or of graduation, have stood high on any criterion of competence or achievement that could be mustered. Other institutions have drawn on applicant pools in an area that has provided a hierarchy of institutions for a variety of needs or ability levels; the best example of this is probably the university system in California, where the university, the state colleges, and the junior colleges have aimed at different subgroups of high school graduates, defining these subgroups in terms of level of high school performance (the grade average or rank in class). In most areas the rise of normal schools or teachers' colleges to prepare teachers during the public school boom in the early part of the present century provided a second set of institutions that, because of the circumstances, had to attract and service a different level of student from those already involved in the major state universities or private institutions. But whatever the factors, by 1950 (and indeed the same holds true today) there was probably no instance where within the boundaries of a reasonably heavy population area the best student (in terms, say, of SAT score) at one college would not be in the bottom quarter of the student population of another college equally accessible geographically. In these cases, it may be that these institutions were not organized on a do-or-die basis (per Harvard-like standards) so that higher education became available to virtually everyone who could complete secondary school. That such an educational system has paid off is evident from the societal roles the graduates of these institutions have played.

Still another characteristic way that qualitative standards were maintained is indicated clearly by the classic Pennsylvania study of Learned and Wood.[14] In almost any institution of any size or complexity there has been a tendency for different kinds of students (in terms of level of ability) to be attracted to different majors or programs of study. This intra-institutional diversity, in large part a function of the substance of different fields of study (e.g., business management vs. theoretical physics), allowed many students to rise or gravitate to a field and level appropriate to the college departmental standards.

The most important and influential selection activity of these colleges is still to be identified: that is, the selection that took place at the hands of the institution through the administration of its own requirements for satisfactory performance, or through the course-by-course performance requirements maintained by the faculty. Although the two qualities cannot be compared, the range of diversity in level of attrition was probably as great as the range of diversity in levels of entering students. At the midpoint of the century, some institutions failed to graduate, for academic reasons, less than 5 percent of those students who entered as freshmen; for others, academic attrition rates ran as high as 80 percent. To my knowledge, no one has made a thorough study of the factors associated with these differential attrition rates; in general, those colleges with a high quality of entering freshmen have tended to fail smaller proportions, but there were some important exceptions. As in the case of the Board-type colleges, applicant supply and demand, itself a partial function of institution supply and demand, has probably been the most crucial factor.

What, then, did the first half of the present century add to the theory and practice of selection for college? Clearly, the most important developments were those associated with the concept of mental ability, and the perfection of techniques for measuring it. Rather than viewing suitability for college solely as a function of the content and quality of the student's preparation, those responsible for admissions practices began also to look at the student's "sheer power of mind." The focus shifted from the school and curriculum to the student himself. Reasoning tells us this could not have come about without some workable levels of quality being attained generally in the secondary schools; but other factors contributing to the use of the mental ability criterion were developments in the field of psychological measurement, needs for greater measurement efficiencies as the numbers of students to be measured increased, the catalytic impact of World War II, and perhaps the realization that the difficulties in legislating quality and content in the high schools by the college through selection standards were, after all, insurmountable.

This development shifted the burden, if not the responsibility, for defining admissions criteria to the test technician. To survive, the technician could not operate simply by his own whims and fancies; he had to take cognizance of the interests of those he served. Research was also part of his way of life, and much that had been taken for granted was subjected to painstaking analysis and scrutiny. Although then, as now, there were vocal critics among the academicians who feared gross inadequacies in the new tests, the technician found that in practice he was taken more seriously than he desired to be, and much of his time had to be devoted to preventing too much faith in his product by the consumers.

The advent of the measurement specialist, as we shall now call the technician, made selection a subject of scientific inquiry rather than a matter of academic debate. For the measurement specialist, the most appropriate model for this inquiry was regression analysis, or prediction. Thus, a corollary of the shift to the mental ability criterion was the formal examination of how well any admissions criterion, however defined, predicted later performance in college. This procedure provided a way to take into account the aggregate judgment of the members of a college faculty by relying on their later evaluation of student performance as the criterion for validating the new tests.

Other interests the measurement specialist brought with him were his emphasis on reliability of student evaluation procedures, and his proclivity for the "objective" format; subject matter achievement tests still seemed reasonable for the task of predicting later performance; the colleges demanded them, and the measurement specialist found that he could make a contribution here as well through placing the examination questions in the "objective" format.

Another aspect of developments in the first half of the twentieth century was the recognition, in practice, that the secondary school both maintained and merited a considerable say in the question of who should go to college. This, in turn, focused attention on selection for higher education as a product of "a series of selections."[15] In a forthcoming paper, Dyer describes the admissions process in the following terms:

> It is a process that has large consequences for the careers of individuals and for the character of the society of which they are a part. The selections may be deliberate decisions by parents, students, and institutions, or they may be the result of social, cultural, and economic forces outside the range of individual human choice. The manner in which an admission system operates is thus partly a product and partly a determinant of these forces.[16]

The role of the elementary and secondary school in the selection process was further emphasized by the great mass of U.S. colleges coming to depend heavily on the simple fact of graduation from secondary school as evidence of eligibility for admission to college. The colleges found, however, that they were not hog-tied to a common, homogeneous mediocrity, but that forces (some within the college and some outside) other than those controlled by the institution through its admissions procedures could be used effectively to permit individual institutional freedom to define goals and standards. Nevertheless, the period showed that student supply and demand was a powerful force with which to contend.

Somewhat at odds with the foregoing, there began to emerge a status hierarchy that seemed to derive from selectivity at the point of admission to college itself. It was the most reputable and distinguished colleges, those blessed with the most esteem by the academicians as well as the general public and with the most applicants, that employed the test systems for excluding applicants. Selectivity beyond the minimal dependence on completion of secondary studies came to be regarded as virtually synonymous with quality of institution. That quality of the entering student was related to the quality of the institution could also be rationalized by the reasoning that the more able the student, the more rigorous and advanced college work he could sustain. Ergo, as supply factors permitted open-door colleges to become selective, they tended to adopt the procedures of the selective colleges, not only because they were manageable and available to imitate, but also because these procedures seemed to promise an attractive qualitative evolution for the institution.

Finally, the period surveyed saw not only the emergence of the measurement specialist, but also the turning of attention of those more generally concerned with college admissions to the management of the specialist's products and to the broader societal problems this responsibility dictated. The College Board, representing the interests not of one institution but those of many institutions, could now be free to look more closely at national educational interests.

## The Prediction of Success in College

It would take several volumes to survey adequately, study by study or even study-category by study-category, the published research concerned with the prediction of success in college; and unpublished studies probably outnumber published studies by a ratio of about fifty to one. Before attempting to place a few of these studies in some historical hierarchy or before moving to major themes in the most promising modern work, it would not be amiss to speculate on the reasons for all this activity.

It has already been noted that as the voice of the measurement specialist was heard in our land the time of formal research in selection had come. The measurement specialist could not stand, as did the academician, on his definition of the essential components of an area to be tested; virtually everything he did had to be verified by placing the measure against an externally defined criterion. Whether building a test of achievement, or of some kind of ability, behavior, or motivational construct postulated to be associated with academic performance, the measurement specialist's regimen demanded formal validation and cross-validation studies.

A second reason, which drew its power from the teaming up of the measurement specialist, the subject matter specialist, and the admissions officer, was the need to determine how well the measures worked for each institution, and to prescribe practice from the guidelines provided by the validation research. There has been some variety in the student populations, predictive components, and curricula, thus frequently making these studies of potential general interest. Also, as selection is practiced for a time, the range of the admitted portion of the population on the predictor variables tends to reduce, making repeated studies appear important.

A third reason for this activity has been the difficulty in capturing more than 25 or 30 percent of the criterion variance in the nets of the predictors. It was the measurement specialist and his experience that determined what a reputable level of relationship between ability or past performance should be; what he achieved appeared useful. But substantial unexplained variance remained; some individual students conspicuously performed much better or worse than the predictive indices indicated might be expected. The validity barrier, like the sound barrier, has been a limit which men fain would push beyond.

A fourth reason has been the infinite versatility of all of us, measurement specialist, admissions officer, teacher, clinician, or simply parent with child in school, to postulate a variety of traits or characteristics that would appear to be associated with academic performance. It would seem patently clear that factors other than native ability and prior achievement affect academic performance: study efficiency, motivation, intellectual curiosity, freedom from extraneous concerns, special interests, traits of character, *et al*., *ad infinitum*. Sometimes it has been the psychologist who believes he is zeroing in on some basic and fundamental moderator of behavior; sometimes it has been the layman who wishes to show the psychologist a thing or two. The triumph of certain insight can sustain the crowning of empirical verification, and so the studies have been conducted--and sometimes reported.

A final reason, but not the least significant, may be that validation research in this area seems not difficult or inconvenient to conduct. Applicants or students can be coaxed or coerced to submit themselves to an examiner, their later academic performance may be copied from the records, and the investigator can then look at mean performance of subgroups at different levels on his predictor measures. Given a brief excursion into the jungle of elementary statistics, he may discover that he can become the grandest tiger by exercising himself through correlational analyses and tests of significance. To the resulting confrontation, the Little Black Sambos who edit professional journals have not infrequently yielded a block of pages.

All of these elements are apparent in the first major published reviews of the literature of prediction of academic success. Harris found, in 1931 and in 1940, a large number of studies and a considerable variety of experimental predictors.[17] Many predictors beyond intelligence and achievement ones worked well, but only when intelligence was ignored or not controlled in the designs. For the times, this is understandable; Brigham, Terman, Toops, Otis, and a horde of other now notables had translated the notion of Binet to successful outcomes; others now wish to try the same from Watsonian or Freudian bases. But if these other measures proved valid, the best explanation was always because of the mutual dependence of the predictor and criterion on the underlying cognitive factor best measured by tests of mental ability. Harris concluded that methodologically we were a pretty sorry lot, and argued that we should learn to profit from the mistakes of others, if we could not contrive better designs of our own.

The next major and competent published review was that by Fishman and Pasanella,[18] covering the period from 1948 to 1957. Fishman's summary of that review in a subsequent paper speaks concisely of what they found:

> It would hardly seem to be too much of an exaggeration to say that nearly every investigator of higher education has done a study predicting college achievement or adjustment. It also seems that every investigator has done only one such study.
>
> What is the upshot of all this research on college selection and guidance? Unfortunately, it can all be summarized rather briefly. The most usual predictors are high school grades and scores on a standardized measure of scholastic aptitude. The usual criterion is the freshman

> average. The average multiple correlation obtained when aiming the usual predictors at the usual criterion is approximately .55. The gain in the multiple correlation upon adding a personality test score to one or both of the usual predictors, holding the criterion constant, is usually less than .05.[19]

The failures to improve much on prediction over scholastic aptitude and achievement measures can hardly be attributed to absence of ingenuity of the experimenters in seeking or contriving a variety of new potential predictors. First of all, it would seem that all psychological tests appropriate for this age group, devised for any purpose, have been tried a few dozen (sometimes a few hundred) times as a predictor of performance in college. This applies particularly to the mass of personality inventories and the interest tests. Of course, the majority of these tests were contrived for other purposes. But further, scarcely any concept that would appear to be pertinent to academic performance, efficiency, or satisfaction has not been transmuted into a measuring device of some sort. We have tried tests of study habits and attitudes, achievement motivation, reading efficiency, and the like. The physiological indices have not been ignored; here experimental predictors have ranged from rate of growth of testees to vitamin (vs. placebo, of course) supplements. A great variety of biographical or personal history items have also been explored, together with factors drawn from the nature of previous academic experience (e.g., public vs. private school background, or special educational treatments). When one excludes those studies that do not provide controls for intelligence and past achievement, the occasional successes are almost always matched, on replication attempts by the original investigator or by another, with failures to confirm the previous findings.

The serious student of prediction of academic success should not accept this report of failure and seek some more hopeful profession without examining this literature in greater detail than is possible here. More optimistic modern reviews are those provided by Stein[20] or Lavin.[21] These and the earlier literature do provide some small pockets of hope.

One of these hopeful signs may be drawn from the occasional consistencies in the findings that are positive, if of small degree. Both the studies of biographical factors and of social and demographic variables point to some pragmatic underpinnings of academic success or persistence: attitudes toward school, having a vocational goal (males only), parental educational level (or perhaps socioeconomic status). A consistent superiority of public school students over private or military school graduates of equal ability has been found (e.g., Koos, in 1931;[22] Davis and Frederiksen, in 1955;[23] Shuey, in 1956[24]); the most plausible explanation of this phenomenon is probably that advanced by McArthur in 1954 and 1960,[25] who suggests that in American society achievement motivation and upward mobility through educational attainment is essentially a middle-class and not an upper-class value orientation. If any consistent thread runs through the studies utilizing personality tests or inventories, it is academic orientation or achievement motivation.

Yet the relationships that have been discovered are small, and for each consistency several new challenges present themselves. One of the most sobering of these is the sort raised by a number of investigators (e.g., Holland, 1959, 1960;[26] Getzels, 1960;[27] Davis, 1965[28]) who, in examining the criterion of performance provided by instructors' grading, have found evidence that the cooperative student, the one who conforms to the value systems of the teacher, or the convergent (as opposed to the divergent) thinker is the one who may have the distinct advantage when performance is evaluated. Getzels, who sees the conventional tests of ability and achievement as well as the conventional criteria stacked toward "selecting the college student mechanically as Manpower," states the problem: "It is the convergent individual who is the most ready source of manpower, the divergent individual the best hope of Man."[29] (He concludes, meekly, that we need both.)

Another sobering challenge, beyond the notion of inadequacy of the criterion, is that provided by the question of whether we are trying to force on selection some of the burden we should be shouldering in our teaching and academic goals. Were there a good measure of achievement motivation, should we select students for college on that basis, or should we use it to determine what instructional activities or contrived educational experiences promote it? One may argue, as have Coleman and Cureton[30] who found statistical justification for the notion, that intelligence is associated with achievement because, in effect, the teacher's tests or other grading practices provide mostly a home-made (and perhaps somewhat specialized) second test of intelligence. If taken to extremes, this can be a terribly damning indictment of testing for selection to college and for what goes on in college. Are we insuring, by selection, that whatever we do or fail to do by the curriculum and our teaching, our students will emerge only with an underlying innate or long-before acquired facility to memorize, to handle abstractions, to manipulate symbols? In grading, do we certify their original promise instead of our impact by repeating the predictor measure?

We could probably feel more comfortable in these arguments if we had devoted the same attention to the criterion that we have given the predictors. In one sense, the earlier concern with predictors would seem justified because it would seem better that the measurement specialist rely on competent outside authority to determine what is good or useful. But then there are nagging studies such as those by Page[31] on essay grading by computer; he has found that a computer, primed to detect common misspellings, or to count unusual words, number of dashes, or number of words, can predict the composite evaluation by trained judges of general writing ability better than any single judge can predict the consensus of his colleagues. Similarly, the studies of Klein and Skager,[32] through focusing on the criterion of expert evaluation of esthetic products (art sketches), have revealed some simple, mechanical guidelines that can be taught to secretaries in a " five-minute art appreciation course" and which enable them to match the evaluation of the professionals. In other words, the attention of the measurement specialist to the criterion begins to seem justified: not merely to prove, after all, that his predictor was reasonable, but to help the expert realize the limitations and sources of bias in his own evaluation as an initial step in contriving better criteria. At any rate, the measurement specialists are beginning to show that they have a contribution to make on the other axis of the regression plot, and are gathering the courage, or the security, or the capability required to tackle it.

There are some other general lessons in the literature of research on prediction of success in college. I have noted in other reviews several instances where the employment of the concept of moderator variable has permitted reasonable improvement of prediction.[33] The classic work here was done by Saunders[34] and Frederiksen and

Melville,[35] who found that appropriate interest scales of the Strong Vocational Interest Blank predicted academic success for non-compulsive students but not for compulsive students; before treating the data in terms of compulsiveness (the moderator variable), no relationship was apparent. Ghiselli[36] has also devoted attention to this technique. The possibilities here are staggering, for one can, starting with sex, postulate moderator or moderator-of-moderator variables by the hundreds.

Another lesson has been that if trying one new predictor after another is not likely to work, or if diverting a new clinical instrument to the purpose of academic prediction is a frustrating exercise, then long-term persistent efforts by a competent team may be the answer. The work of French and his associates[37] on differential prediction of grades in college involved painstaking item construction for a battery of cognitive variables, careful identification of separate sources of variance through factor analysis, and, incidentally, an interest test constructed specifically for the academic prediction task. The place of the cognitive factors is not yet assured, but French did find the interest test adding to the prediction of grades and predicting satisfaction with courses as well. The work of Schlesser and Finger[38] with the Personal Values Inventory is another case in point where involvement over time with a tailored rather than adapted instrument seems to be paying off (although they have persisted in failing to control for past achievement in secondary school).

There is also evidence that raw empiricism, or the blind search for expedients that work, is not enough; Messick, our country's young dean of personality research toward academic prediction purposes, has stated the case in a significant paper that should be read in the original.[39] His arguments and work attest the need for careful preliminary theorizing or model building, and for exhaustive construct rather than predictive validity research.

Another argument too well thought out to be ignored is that of Fishman,[40] who argues for a social-psychological rather than a simple regression model. He makes a nice case, for example, for looking at the similarities and the differences in the tasks and values systems or environmental characteristics of the secondary school vs. the college, in understanding the successes and the failures in predicting success in one setting from success in the other.

Some exciting possibilities from the current work will be saved for the second part of this paper, where, after examining briefly other applications of measurement science, I shall attempt to speculate on the most promising leads for future study over the variety of potential applications. This is appropriate, if we accept as our criterion the fact that current selection practice has not yet added universally accepted new measures or practices. We have seen the reasons for this failure as stemming from too opportunistic a flitting from one new measure to another, too much blind empiricism, too little attention to (or too ready acceptance of) the criterion to be predicted, failure to take the complexities of behavior into account through the moderator variable approach, too much reliance on the regression model to the exclusion of new models, and failure to accept the necessity of long, massive, painstaking work if success is to be achieved. It is accurate to say that no one, whether some measurement genius working in a university office or laboratory as did Brigham, or the several hundred specialists at ETS with its expenditures for research and development of more than $2 million last year, has any promising new mousetrap ready for our market or for import.

Before leaving this area, it is imperative to crystallize one more implication of the body of research: that is the need, stated most effectively by Messick,[41] to consider the ethical bases of selection. If we find that Jewish applicants outperform non-Jewish applicants, shall we be preferential? If young people from broken homes appear less likely to persist in college, shall we discriminate against them? If our current tests and criteria are biased against culturally different subgroups of the population, are we justified in continuing them? Many of the personality variables tried are even more problematical, for they imply not only unreasoned bias for particular qualities but also abdication of the institution's responsibility for changing behavior. Applicants of considerable closed-mindedness may seem less desirable and be deucedly difficult to teach; the John Birch Society or the Ku Klux Klan may accept them if Siwash U. considers them undesirable. But in another perspective these applicants could be considered more as evidence of failures of the educational system gone before than as undesirables for further training. Many qualities of personality are, of course, the product of extra-school forces; and there is not widespread sympathy now in America for the assuming, by the college, of a therapeutic role (to the point that we are experiencing a closing or phasing out of the many university counseling centers that sprang up during the war years). The writer stands strongly with those of the opinion that the function of the college is to teach, not to treat; but in many areas of personality development, good teaching in the traditional disciplines may be as effective as the psychiatrist's couch, barring serious psychosis, and many of our most cherished academic goals have liberalizing or personal style components. It would seem we are on dangerous grounds if we rush too quickly to establish sets of good vs. bad personality traits for use in selection. The final resolution may have to involve not only the measurement specialist, the subject-matter specialist, and the admissions officer, but also the teaching technologist, the humanist (particularly the philosopher), and the significant leaders in the architecture of educational systems.

## The Colleges and Admissions Today

If we preserve the continuity in the historical tracing of the development of American higher education, its admissions practices and its testing movement, we must turn now to the period from about 1950 to the present. There are many social, cultural, and educational factors that have influenced higher education and selection in the last two decades: the information explosion, the rapid development of educational technologies, the ubiquity of the computer, the civil-rights revolution, the spread of urbanization, the legitimizing of the college as a focus for research by social scientists. But none of these forces has had the profound and direct bearing on the institutions of higher education that increased enrollment pressures have caused.

The enrollment pressures are, of course, a function of the higher birth rates during (and continuing after) World War II, as well as of the larger proportions of the population seeking higher education. These larger proportions are, in turn, derived from the continuing expansion of belief in education, a national prosperity that permits later entry into work as well as bringing higher education into financial reach of more people, and the fact that in America old institutions expand and new ones spring up if there is a demand from prospective applicants.

The best summary of the extent of this increase in the United

States is probably that given by Dyer in his forthcoming analysis of college and university admissions.[42] First, drawing from Bowles[43] and U.S. Bureau of Census data for 1960, he compares "the severity of educational selection in the world at large with its severity in the American system:"

Table 1*

World and U. S. Enrollments as Percentages of Age Groups

| Educational level | World % 1950 | U.S. % 1950 | World % 1959 | U.S. % 1960 |
|---|---|---|---|---|
| Primary | 37 | 93 | 50 | 95 |
| Secondary | 18 | 75 | 27 | 85 |
| Higher | 3 | 15 | 5 | 24 |

*Table taken from Dyer MS.

Dyer also notes, from various data:

> Of the total number of American children who entered the fifth grade in 1956, 36 percent entered college in 1964 by normal progression through the system.... Of the cohort that started the fifth grade in 1942, only 21 percent entered college when their time came in 1950.... In 1964, 54 percent of the high school graduates went directly on to some form of college work as compared with only 41 percent of the earlier group. This rate probably reflects, to some extent at least, both the greater pulling power of higher education and greater accessibility of its institutions.... Since 1950 the college admissions curve has become steadily steeper:
>
> 517,000 entered in 1950,
> 690,000 entered in 1955,
> 929,800 entered in 1960,
> 1,453,000 entered in 1965,

and conservative projections say that the number of new freshmen in 1975 will be something over 3,000,000.[44]

Another factor making this possible, even reasonable, is the already noted diversity that has emerged in American higher education. This has been documented by a number of people in various ways.[45] It has been frequently noted in our country in the past fifteen years that there is a college for every high school graduate, whatever his ability level.

Yet the demand for admission has not been evenly distributed over the range of institutions. Some are more popularly aspired to than others; some tend to draw from certain segments of the population in terms of ability level or socioeconomic status; some try to expand (generally the public institutions), while others try to maintain their size (generally the elite private institutions) but exercise greater selectivity. Although some colleges have turned away as many as nine out of every ten applicants, others have needed additional students to fill classrooms and dormitories.

Perhaps, as I once noted,[46] selectivity would seem a rather powerful way to manipulate the character of the institution in desirable directions. Perhaps selectivity is a distinguishing characteristic of the distinguished colleges and universities. Ebla's volume[47] leaves little doubt that a status hierarchy related to selectivity exists among

U.S. institutions and goes beyond athletic rivalries. This hierarchy does not exist solely in the minds of the scholars and academicians, but is shared in general by the American public, with the result that the tightening of selection procedures in a college sets off a geometrical progression of greater and greater selectivity. With enrollment pressures increasing, the most selective colleges have become of necessity even more selective. A great many of the established public universities, and many other private colleges with relatively fixed endowment, have suddenly had the luxury of an abundance of applicants and have enjoyed some additional freedom not to try to squeeze all aspirants inside by the development of newer institutions--notably the junior or community colleges, now being created at a rate greater than one a week on the average--or by the willingness of many of the lower institutions on the status totem pole to try first for greater capability through expansion.

For the institutions that have long been selective, putting prevalent and uniform focus for selectivity on one standard cognitive criterion, there has been a running out of range on this measure. When a university enjoys more valedictorians and more high scorers on the SAT among its applicants than it can accept, how then is it to sort? This problem has had two different kinds of impact. The first is to make the search for additional bases for selection seem more urgent; the second has been a return to more comprehensive achievement testing. On the latter matter, the current Executive Vice President of the College Board has recently commented:

> While the number of candidates taking the SAT has grown two and a half times since 1959, Achievement Test volume has better than tripled. Whereas the decade of the 1950's might be termed "the aptitude era," the first five years of the 1960's could be called "the return to achievement." The most dramatic demonstration of this trend has been in the Advanced Placement Program, where the numbers have increased fivefold in five years. A throwback in a sense to the syllabus-based comprehensive tests of the 1930's, the Advanced Placement Examinations have served to reimpose restraints on the secondary school curriculum directly through the course descriptions for them and indirectly through the prerequisite to advanced placement work.[48]

For the mid-range, traditionally "open-door" institutions, the enrollment pressures have brought into being the possibility that they may both grow and begin to experience some selectivity too. It was this market to which--particularly in the Midwest where the College Board seemed more like a New England property--the American College Testing Program seemed to appeal, at least to the extent that a husky second national admissions testing program grew into being in the latter half of the 1950 decade. In 1957 the Regents of the University System of Georgia enacted the bold requirement that all applicants for any of the seventeen public colleges and universities in that state would submit scores on the Scholastic Aptitude Test of the College Board; these scores would be used initially in a Regent-sponsored program of research, but the hope was that the stronger colleges academically could be more restrictive and more efficient (through taking fewer higher-risk applicants), while the others could accommodate those who would profit from less rigorous studies or different programs, or serve as a proving ground for those for whom it seemed unsafe to start in the bigger league. It is the movement of the mid-range institution into the prospect of selection, rather than increased applicants for institutions which have long been selective, that has accounted for the boom in admissions testing (from a volume of admissions program tests of 746,522 in 1959-1960 to a volume of 2,076,470 in 1964-1965, according to the annual reports for those years of Educational Testing Service; for the younger American College Testing Program, the comparable volume figure for the latter year was 705,063).

For the new colleges and those that would remain relatively open, the selection testing movement has taken another relevant turn. It is precisely these institutions, with their wider range of applicants, or with their greater preponderance of students of lower academic potential in the traditional sense, for which the tests would seem most usefully discriminating among students. These institutions are also frequently infested with a variety of programs. Here, too, testing with the examinations developed for selection has begun to build up, with the emphasis on pre-college counseling or guidance, determination of specific needs for remedial work upon entrance, or placement. In this spirit the College Board has produced the

Manual of Freshman Class Profiles for Indiana Colleges (CEEB, 1965). It is significant too that the Board is currently mounting a major effort to develop a useful battery of tests for community colleges.

To look at present times solely in terms of selection test practices is to miss one important flavor in admission to college in America. This flavor comes from the colleges and their admissions officers. Although these are the consumers of selection tests, it would be a gross error to imply that test- or achievement-based criteria are their sole concern. From a review of college catalogues and statements of admissions officers, I have summarized their concerns in the following way:

> Our question now is: what qualities do admissions officers seek? Several classes of criteria can be found from a review of such statements, or from studying the admissions process in a number of institutions.
>
> The most frequently cited class of criterion is that which pertains to qualities directly applicable to academic achievement. Ability tests and past achievement are the main ingredients, but some colleges extend with these names (at least) of unusual talents or such characteristics as "thirst for knowledge." In this class of criteria are all those personal attributes that academic man has postulated in himself and his prized students.
>
> A second class of criteria has to do with traits or qualities generally valued in our culture. These have little or nothing to do with the business of learning, but reflect the opinion of important constituencies of the college as to what constitutes glowing young manhood or womanhood. Typical are qualities of "Christian commitment" in church-related colleges, "leadership" in colleges envious to have a place in the sun through their graduates, or "personal stability" in colleges without facilities or courage to deal

> with troubled students.
>
> A third class of criteria has to do with practical advantage or necessity, or specific needs for maintenance of the institution and its particular programs. Colleges with affluent alumni admit preferences for alumni sons; athletic programs are maintained; state quotas are filled first at public colleges. These can work in reverse: one prestigious private college has found a preponderance of its applicants Jewish, and has found it must restrict here more severely or lose its traditional secular character.
>
> A fourth class of criteria has to do with the hope of obtaining a balance of students from identifiable and hopefully meaningful subgroups. Here, the search is for variety that is meaningful in itself. The provincial college anxious to become more catholic may give priority to distant applicants; a college with many urban applicants may seek, as Harvard has stated it does (Glimp and Whitla, 1964), applicants from a rural or small town background.
>
> Some colleges have been bold (or confused) enough to call for a still more sophisticated variety of students--this time, in terms of a catalytic mix of personal qualities. Extroverts may be balanced against introverts, four-letter men against shy, young, thick-lensed scholars of fourteenth century French poetry, and so on. The student body is certainly a potent source of stimulation; whether it can be manipulated effectively this way remains to be seen. Yet, some students were admitted last year in selective colleges primarily because they added some sort of seasoning to the freshman broth.[49]

These criteria indicate, first of all, that American colleges believe fiercely that they have the right to select their own students. But, of the classes of criteria, measurement specialists or other educational research personnel have researched only the first, that pertaining to qualities related to academic achievement. Research in the other areas may be difficult, but their content indicates a need for widening the scope of our studies.

Our final area for review of current thinking and practice in the United States draws from the burden of research, the trends in higher education, the concerns of admissions officers, and the new areas of involvement of the College Entrance Examination Board. This has to do with the examination of selection broadly in terms of its societal implications, rather than what it seems to promise test by test or institution by institution.

There is no longer much vocal concern about legislating quality and content of the secondary school through selection standards. Selection may seem a simple way to enhance the quality of an institution of higher education (this has probably been so for many years), but that is now seldom disguised as an attempt to impose college standards on the preparatory agents.

In the broader sense, selection is more frequently viewed as a system (affected by many forces appearing from the time of conception to the time of entry into work) for exercising our biases for the control of society. With higher education becoming more pervasive for our population, the question has become less that of "Who shall be educated?" and more that of "How and where shall various individuals be educated?" This view has major manpower implications, and means that the study of selection should have one foot firmly on the channels of access through which people pass, and the other on our manpower needs for people with various kinds of training.

As our attention has begun to shift from the sorting that takes place at the point of admissions to the sorting that takes place at other points and by other forces, we have learned that self-selection--the selection of the college by the applicant--may have been more influential all along. A more reasonable control than conscious, deliberate sorting among applicants for admissions may be that which would augment pre-college guidance at the pragmatic end, and explore college image or other factors that attract and distract different kinds of students at the developmental end. These are goals of the Board's current guidance program; and the Board has commissioned a major longitudinal study of channels of access (a report of pilot work has been published by Trent)[50] that is now under way at the Center for Research and Development in Higher Education at the University of California at Berkeley.

Last, but not least, is the emerging role of the measurement specialist. It would seem that he can no longer afford to content himself with the construction of predictors of human behavior, but he must also enter the business of helping define the criteria by which that behavior and his predictors are evaluated, and the characteristics of the institutions, teachers, and personal experiences in which new behavior patterns are acquired. In many cases, he is finding that his knowledge of his predictors, and the constructs on which they are based, are useful in diagnosing the components of the criterion and the situations that produce people who can meet it. Although his knowledge of his own limitations has grown, he has also in that process acquired some cogency among his own cult, and now, if he is good, has designs on attempting to facilitate the impact of the educational system on the individual and on the society of which that individual is a part.

# II. APPLICATIONS OF MEASUREMENT IN HIGHER EDUCATION FOR THE SECOND HALF OF THE TWENTIETH CENTURY: ACHIEVEMENT AND PROSPECT

## The Emergence of a "Science of Measurement"

### The first half of the twentieth century

The turn of the present century saw not only the formation of the College Entrance Examination Board, a multi-institution agent that could become a major consumer and purveyor of the output of measurement science, but also a general turning of attention to two new areas of inquiry that were to form the crucial basis for a science of measurement. One of these areas of inquiry grew out of concerns within the discipline of psychology that individual differences must be considered, that a search for modal or general laws of behavior was not enough. Thus, the search was on for meaningful human traits or attributes in which individuals varied, for ways in which these differences might be quantified, and for practical implications of particular human variability. The tenor of the times just before 1900 can be captured by visualizing Professor J. McKeen Cattell at Columbia University asking his class to stand with arms outstretched as long as they could, scoring this little test with the click of a stopwatch as weary muscles gave way and arms dropped, and then seeking a relationship between scores and academic performance (the results of this pioneer attempt to assess general motivation were negative).

Shortly after 1900 the French physician Alfred Binet, who had been concerned with the diagnosis of mental defectives and with how "bright" and "dull" children differed, found an opportunity in the Paris schools to attempt, through testing, a sorting of children into regular or simplified programs. This first real test of intelligence or general mental ability proved both successful and useful, and found quick translation (in 1910-1916) and further development in the United States principally through the efforts of Professor Lewis Terman at Stanford University, who reflected the growing interest among American psychologists in the general psychological development of the individual.

The second major area of inquiry needed for the development of a science of measurement was that of statistical procedures for dealing precisely with arrays of measures. Sir Francis Galton, his interest aroused by Darwin's new theory of differences among species, was not only concerned with the invention of ways to measure physical characteristics, keenness of the senses, and mental imagery, but also mathematical ways of expressing differences among individuals--of placing the person within a group, and of describing concisely the group as a standard for comparison. Galton and other Britishers such as W. S. Gosset (a statistician for the Guiness Brewery) carried such concerns on to various extensions of the two grand concepts in statistics: that of procedures for assessing the degree of relationship between concomitant measures or conditions, and that of probability and significance, where (in simplest terms) one is concerned with whether observed differences in measures are the result of error in measurement or of some more basic underlying condition.

Although in the other mainstream there have been occasional geniuses who have made a significant if lonely mark, trait definition and development of test or measurement devices has been, oddly enough, a product of hard times in a generally affluent and literate society. American psychologists have clearly dominated the field in both the invention and the application of measures (Great Britain is just now, in the late 1960's, getting around to determining, through its Vice-Chancellors' Committee on Entrance Procedures, if the objective American admissions tests have any relevance for them). Also, history shows that the test development milestones in significant and successful production of human measuring devices have come in periods when candidates were either widely available or there was a need to select them quickly (both of which place a premium on choosing without waiting for a test of performance over time). Thus, the first boom in the construction of measuring devices came with World War I, when groups principally at the Carnegie Institute of Technology concerned themselves with the invention of measures for classification of servicemen toward assignment to training opportunities.

Psychologists involved in that effort, such as Edward K. Strong, Jr., Herbert Toops, or Arthur Otis, were to hold their place in the psychometric sun for the next several decades.

We have already cited the adaptation of these efforts to the college admissions problem by Carl Brigham, the fruits of whose labors in the 1920's led to the modern College Board tests. A landmark study with implications for higher education was a sweeping survey by Learned and Wood of student achievement throughout the State of Pennsylvania in both lower and higher institutions of education.[31] In the late 1920's, these investigators (in a mammoth project supported by the Carnegie Fund for the Advancement of Teaching) looked over many institutions and departments with before-and-after measures of academic achievement. The findings obviously took the investigator's breath away momentarily. In the main, far more variation in level of growth than had been anticipated was found both among institutions and among departments within those institutions. (Indeed, college sophomores in some departments were found to rank with tenth-grade students in the general population!) Needless to say, the findings were as threatening as they were informative; and today, with increased knowledge of the limitations of tests and of their reliance on the curriculum, a less dramatic case would be made than that of the investigators in their report. But our first finding of great significance for a measurement science with implications for higher education is that considerable diversity exists among and within institutions in the level of input and output on student achievement.

One might suspect that after such findings had been reported, other states, or collections of institutions, or guidance interests, would rush to chart the levels of diversity among meaningful groups of institutions. But such was not the case. It may have been some part of the American dream that holds a college is a college is a college; it may have been that institutions do not willingly submit to the prospect of a ranking, or that administrators are concerned with not tipping internal or external balances of confident complacency; it may have been that officials of multi-institutional agencies with responsibility over many institutions did not read the findings, or, if they did, were afraid of threatening the inevitable half of the constituent institutions that would fall below the average in any ranking; or, it may have been that colleges in the depression years were so struggling for survival that those students with funds for tuition were happily accepted, whatever their credentials. But the next significant attempt to openly describe a collection of institutions probably did not come until some 20 years after the publication of the Pennsylvania study. In 1957 J. A. Davis, from the protected position as staff member for the Board of Regents for the University System of Georgia, published a "Counselors' Guide,"[52] giving test data on entering freshmen in the seventeen-college system of that state.

The most significant effort at developing and expanding the store of measures of human traits and attributes came in work at the Minnesota Employment Stabilization Research Institute[53] in the 1930's. This effort was not concerned particularly with measures with implications for higher education, but rather with generally lower-level vocational aptitudes that had relevance for vocational success. It did provide some experience in extending the test-contained " samplings of behavior" to a variety of new areas, laying a basis for a later search for constellations of unitary " aptitudes" ; and it did begin to signal the complexity of human traits and the lack of preciseness possible in defining the minimal levels necessary for an individual's successful performance in some societal role.

Shortly before this time, Spearman in England and Thurstone in the United States began to examine conventional mental ability testing to determine if the variance these tests captured could be divided into separate, unrelated traits, and if our concepts of mental ability represented a conglomeration of things which were worthwhile but of differential applicability. The tool they used in this process was an extension of correlational techniques that came to be known as factor analysis. Although debates raged for several decades as to whether there were general (Spearman) or specific (Thurstone) " factors" making up mental ability, the controversy led more toward perfection of the statistical technique than toward demonstrating that one theory was more nearly correct than the other. Perhaps the most significant aspect of the controversy was that the argument tended to hinge on logical or mathematical proofs, rather than on empirical or utilitarian ones. After a little time that permitted obeisance to such intellectual niceties, the hue and cry began to emerge for attention to criterion measures (those which may be used to validate a predictor measure, or reveal, through experience, what the initial predictor measure was all about) and to empiricism in test construction (wherein the search is for expedients that work, whatever the theoretical basis for defining a trait). With the advent of the computer age in the early 1960's, factor analysis itself was to become a tool for test refinement or clean-up of items, and an indicator of traces of new factors that could be amplified to form new tests.

As the country began to emerge from the depression toward the end of the thirties, there were some extensions of measurement to educational arenas that involved more than the conventional ability and achievement tests. In several instances, new colleges and a desire to innovate carried a social scientist along. This was the case at Bennington,[54] where the research model provided for a hypothesis of attitude change as well as cognitive growth. C. R. Pace[55] picked as a central problem the study of the institution through what could be observed in the subsequent lives of its graduates, as did Chamberlin _et al._,[56] but the measurement aspects of these studies were not highly developed.

As noted in the first part of this paper, World War II provided some flesh on the bones of Professor Brigham's Psychometric Laboratory at Princeton. By the end of the war there were some one hundred test technicians and statisticians in residence, allowing the laboratory, as the Educational Testing Service, to be the agent for the three adopting organizations (the College Board, the American Council on Education, and the Carnegie Foundation for the Advancement of Teaching). That agent, with its experience gained during the war years, provided some measurement devices to what was probably the next landmark event in measurement in higher education: the Cooperative Study of Evaluation in General Education,[57] sponsored by the Carnegie Corporation and the American Council on Education, and directed by Paul Dressel at Michigan State College. This study was notable because it began with elaborate and painstaking efforts to define teaching objectives in general education and in a number of areas therein. It paired test or statistical technicians with subject matter specialists, and proceeded through the specific and tailored construction of measuring devices to sample growth among students toward these teaching objectives. It also applied those measures in a before-and-after fashion with an examination of institutional or instructional differences among or within the nineteen institutions involved.

In retrospect, as we recall the boom in self-studies from accrediting commission activities as well as the study by the American Council of Education, the decade of the fifties can be seen as one in which measurement-oriented people were concerned with institutional objectives specified in some form amenable to observation of student progress toward those objectives. In 1954 B. S. Bloom and his coworkers published a Taxonomy of Educational Objectives,[58] which focused on the general cognitive domain, although this landmark effort has had more direct use in measurement construction in other countries than in the United States. In this country there seemed to emerge an awareness of the gulf between what one might call general objectives of higher education, as specified in the goals section of a college catalogue (and even as purified by the measurement-oriented person), and the now fairly routinized procedure for building a subject-matter achievement test to reflect the highly specific set of tasks implicit in a course unit within a discipline.

Drawing principally on the burgeoning interest in clinical and/or counseling psychology and a resulting interest in mental health set off by Carl Rogers in the forties, a new entrant in the area of measurement was to emerge and develop. This was the test of personality, important in this account for the negative role it was to play in the fifties and sixties in studies of college students. A number of entries, ranging from pathology-oriented instruments such as the Minnesota

Multiphasic Personality Inventory[59] to the normal psychology Guilford-Zimmerman Temperament Survey,[60] were developed with clinical or normal populations of college students or adults, and were to be adapted, as will shortly be seen, to studies of the impact of college.

Thus, in the first half of the twentieth century there had emerged some reasonable experience with a variety of tests of general scholastic aptitude, and some guidelines which made the construction of a reasonably good test of acquisition of factual knowledge in a given area a relatively routine proposition. Application of these tests beyond the college admissions situation, or in multi-college studies, was relatively rare. There were a few attempts to measure by sorting individuals into outcome categories, as in the several follow-up studies. An occasional social psychologist concerned himself with attitudinal change as the function of a particular class of college experience rather than of a laboratory kind of treatment. Some questionable tools for exploring the personal impact of college grew out of the mental hygiene movement. But for the most part, psychologists and statisticians learned how to purify tests toward the prediction of grades in college; colleges learned that these could be useful tools in selective admissions situations, and not much happened beyond that.

### The decade of the sixties

It is not inaccurate to generalize that measurement specialists were, by factors intrinsic in their role, implementers rather than innovators in their first fifty years. This staff rather than line function was to be changed rather sharply by a significant challenge from outside the ranks. The start of a completely new era of application of measurement, and new roles for the measurement scientist, probably began with P. E. Jacob's study,[61] published in 1957, of value change as a function of college experience. His finding that not much change takes place threatened the very core of academia, and many rose to the challenge: some to attack his findings, others to see for themselves.

If Jacob conceived the new mood, Nevitt Sanford (most precisely a personality theorist), through a collection of essays entitled The American College,[62] served as midwife for the birth. Although focusing frequently on relatively narrow personality areas of student development, he also struck provocatively at many of the cherished beliefs and values of higher education. Nourishing this infant was the explosion of the college-age population and a favorable economic situation, both of which encouraged the prospect of rapid change in positive directions for almost all institutions of higher education. With such a start, the expansion of federal and foundation support for the technological revolution in education, and the beginning evolution therefrom, served to make involvement in research not just possible but mandatory. The result is that at least some ten studies, of as much significance as any ten in the entire past history of higher education, have just been published or are in preparation. These include a substantial volume by A. W. Astin[63] on the college environment, provoked by the growing interest in educational climate and based on well-designed cross-institutional studies by the American Council on Education; a well-financed review of the literature on impact of college now in progress at the Survey Research Center of the University of Michigan;[64] and a provocative study of eighteen liberal arts colleges now being concluded by Morris Keeton.[65]

This remarkable acceleration of general interest in researching, through measurement-oriented studies, has involved both the measurement specialist and the general social science faculty, who now use measurement and statistics as routine tools. In effect, it has been a migration of the general social scientist into his own college backyard. Thus, Clark and Trow[66] have focused on student subcultures as other sociologists have examined societal subcultures. Sanford[67] concerned himself with personality development as a function of the college educational experience in much the same manner that other psychologists have sought meaningful relationships between personality and early childhood experiences. On the other hand, there has been a rapid growth of a new phenomenon within the college administration itself: this is the institutional research office or person designated to conduct studies related to the maintenance and further development of the institution.

In a recent paper directed most specifically to this latter group, H. S. Dyer[66] has asked the question: "Can institutional research lead to a science of institutions?" The question, particularly as it comes from so sophisticated an observer, implies that we do not yet have a science of institutions, that the social scientist has not yet achieved it, and that the new institutional researcher who could conceivably contribute to its development may not be able to carry it off. Dyer's argument is that the institutional researcher may achieve a science of institutions, _if_ he can integrate his views with those of the discipline-oriented social scientist (i.e., if he can take on the values, strategies, and tools of the scientist) without losing his focus on the mission of the institution, _and_ if measurement is permitted to play a central role.

To those persons steeped in the liberal art of higher education, these comments may seem patently trivial. Social scientists, reasonable enough (or unreasonable enough) people at faculty meetings, seem impacted with technology and with specificity when they are involved through their discipline with an educational problem. Institutional researchers are also pleasant enough as individuals, but are frequently viewed by faculty educational philosophers as market researchers at best, tools at worst, of the new management mainstream in higher education. Measurement reeks of standardized tests. If these indeed are the forces and the basic ingredient of a new science of institutions, can one really expect much to come of it? Considering the truly basic problems--e.g., contrivance of genuine academic freedom, the best synthesis of the conscience vs. intellect question, manipulation of the individual and the environment to assure a deepening of intellectual awareness, or the future of predominantly Negro colleges--what can one expect to see achieved from the measurement science school or the institutional research school?

The purpose of the second part of this review is to explore the state of the development of a science of institutions of higher education--how far it has come, where it has succeeded, where it has failed; and, what directions are indicated for the future. In this review, Dyer's notion that measurement is central will be taken seriously; this not only helps to define and delimit the area of inquiry, but also may indeed, as Dyer suggests, form the crucial component. For measurement, as Dyer uses it, is not a loose synonym for "test," but the end product of the process of defining some quality of concern with sufficient precision that an investigator (or others) can make more exact comparisons among individuals, groups, or institutions in regard to that quality. In most instances, the measure will have limitations, for it usually involves abstracting parts of a quality that in practice has been fuzzy at best. But the process of defining helps to clarify what the measure is not as well as what it is, at least for operational purposes. This brings some order, control, and openness to inspection by others into the situation. Also, the conversion, through measurement, of a quality to a metric unit permits evaluative comparisons of products measured. The base quality can then be studied systematically in relation to other qualities, and studies can be replicated toward determining how the quality develops and functions. Although in the beginning of a science scope is sacrificed to precision, that precision becomes the base for building.

We also need at this point to define "research," for this title has been given to a wide range of activities from head-counting or gentle speculation from a few observations to massive projects involving elaborate statistical treatment of data. Research is the systematic inquiry into conditions bearing on certain events or outcomes. It starts with one or more specific questions, and it operates through a preconceived and deliberate plan, designed to identify and properly control relevant variables so that their true meaning may be better understood. The deliberate plan involves procedures designed to sharpen or make objective the focus on the variables of concern, and to exclude biases, many of which are likely to be subtle, which may lead to erroneous interpretations or conclusions.

Many who have attempted to support and encourage evaluative research in higher education have attributed such a status to almost any generalization from experience or observation, or to any honest inquiry whether structured or not. Thus, W. H. Cowley[69] sees its beginnings in American higher education in 1701, when Increase Mather, then president of Harvard, functioned as an educational consultant to the founders of Yale. Yet true research must stand on the evidence it marshals rather than on the status of the individual making a pronouncement, the versatility of his argument, or even the eventual proof that he was right. For a long time, many educators bought the argument that training in Latin afforded a unique and valuable mental discipline for the learner. Yet as other subjects applicable to new societal needs began to vie with it in the curriculum, and when true research on the impact of training in Latin was conducted (i.e., when some progress in measuring ability and impact of mental discipline was achieved), it was generally found that the superiority of students with Latin training, or who attended institutions requiring healthy portions of Latin in the curriculum, could more accurately be ascribed to the ability of those attracted to or sorted into courses in Latin than to any habits formed as a consequence of classic studies. Confident, even experienced, opinion can be quite misleading: in higher education it tends to oversimplify (e.g., "good teaching is the crucial force in intellectual development"), or to be too uncritically acceptant of utilitarian or status values (e.g., "a college with high ability entering freshmen is a quality institution").

Even within such a brief historical overview, as this one of the period from 1900 to 1960, it is interesting to attempt to examine the forces at work which may have hampered an earlier development of the research literature that has begun to appear in the current decade. The major deterrent would seem to be the difficulty in acquiring comparable data across a range of institutions. Two factors account for this. First, the fact is that colleges and universities tend to be independent entities, or if parts of a system, seldom have an administrative head who would consider monitoring, through research, the separate components of the system. Second, the very fact of diversity, and the fact that even such an index as admissions test score averages for entering freshmen has terribly threatening overtones for those institutions below the top, indicates that measurement research across institutions involves some very sensitive areas.

The current decade has seen a new force as a most important development toward a measurement-based science of higher education: this is the research organization or team, strategically or by happenstance situated with access across institutions, with ample financial resources for costly inquiries or sustained effort over time, whatever the vagaries of funding, and frequently with some practical operational channel permitting research data as a by-product. The Center for the Study of Higher Education at Berkeley, started in 1957, the Center for Studies in Higher Education at the University of Michigan, established in 1955, and the Institute for the Higher Education at Teachers College, Columbia, started in 1962, are the early pioneers of university-based centers. The regional educational boards, particularly the Western Interstate Council on Higher Education and the Southern Regional Education Board, have provided themselves with a high springboard and have taken a magnificent dive. Among the testing organizations, the National Merit Scholarship Corporation, under Holland's leadership, gave an early and pervasive focus to cross-institutional measurement research. Holland carried this with him when he moved to the American College Testing Program in 1959. Although the College Board and Educational Testing Service, its research and operational arm since 1947, have from their beginnings looked at problems of general interest in the administration of certain functions within institutions, the year 1963 saw the establishment at ETS of a research group concerned broadly with higher education, and in 1964 a new program to provide colleges, at their option, a variety of measurement instruments, packaged research designs, and a variety of data processing services. In 1964 the American Council on Education, a leader for many years in seeing studies initiated, acquired its own in-house research program.

That these giants have become the leaders in one area of concern is no accident. When the chairman of a governing body of a system of colleges, with budgetary control, demands cross-institutional studies, the fearful member institutions have found ways to dodge or defeat the attempt. Cross-institutional studies can, as has been noted, be threatening, for what college does not have a faculty zealous of its own particular independence and brand of control, and a few skeletons in its closets? One needs an organization with easy access to many institutions, a past or a base without the blemish of former crusades, and a financial base to sustain costly activity.

The reader, then, should be aware that in the remainder of this review we are indeed looking at a young plant just ready to sprout; we cannot confidently predict how tall it may grow, or whether it will prove to be prickle or pear. In any event, cultivation may be essential if it is to thrive; or, on the other hand, it may turn out to be a rampant weed that will take some unusual efforts to stamp out.

## The Enlarged Concern with Student Input

If any fact about higher education is well-established it is that a measure of past performance (scholastic achievement), together with a reasonably good test of mental ability (scholastic aptitude), is a good indicator of what grade the student will achieve in college. This finding has held over the years and over the range of institutions, from vocationally oriented junior colleges to selective Ivy League institutions.

It was pointed out in Part I that the first exhaustive reviews of such studies by Harris[70] supported fairly well the conclusions of Fishman and Pasanella[71] two decades later. Performance in secondary school (as attested through rank in class or grade-point average) has been again and again demonstrated as the best predictor of college performance. Adding a test score to this measure improves the predictability, probably because, as J. R. Hills[72] has speculated, it helps correct for differential standards among secondary schools in grading practices as well as because it reflects the basic mental equipment needed to understand and retain academic material. Other measures--of interest, personality, attitude, motivation, or work habits--have not improved the prediction, either because the measures of these qualities that have been contrived are faulty, because these qualities are already subsumed in the measure of past performance, or because these qualities are not so uniformly and critically important as ability and achievement.

Most of the several thousand studies reported in the research literature involve single institutions and the search for the most efficient weightings of the several ingredients of the predictive combination, or the (generally fruitless) search for ways to improve that combination. In the weighting studies, minor variations that make sense have been found: for example, engineering schools find that some minor improvement is effected by weighting mathematical components more heavily than would prove best for liberal arts colleges. But these variations are relatively inconsequential.

Much more substantial are the differences in level (not in content, relationship with grades, or weighting) on indices of academic promise among institutions of higher education. This was apparent in the classic Pennsylvania study of Learned and Wood;[73] the most recent major survey is that by J. G. Darley.[74] No formal and exhaustive studies of why these differences exist are known, although Astin[75] has studied selectivity against a variety of other descriptive indices. His data show that the degree to which an institution must be or can be or wants to be selective, the severity of this selection (a function of number of applicants vs. number of places made available), and the period of time the institution has been selective would appear to be the most powerful influences. Apparently, the practice of exclusion, together with the focus in selection on a rather particular and unitary kind of variable (academic promise from tests and past record), broadcasts to potential applicants some rather powerful signals which are strong enough to encourage some to tune in on other frequencies. The age of the institution also appears to play some part, particularly where the history of the institution extends back more than a century. This may be a function of the relationship between socioeconomic status and intelligence, the fact that in earlier days going to college was confined to narrower bands of ability at the top of the distribution, and some sort of enduring bond between a socioeconomic class group or strata and the institution. Institutional commitment to broad population groups appears also to play a part. Public colleges supported by broad-based tax funds have been generally prone to build to accommodate additional numbers when demand from prospective students among its constituents increases (although there are differences among state universities, whether by age, availability of other kinds of institutions with different selection policies, or from a built-in branching system as in California today). Another powerful influence appears to be the dominant curriculum of the institution, or its designation as an institution for training some particular occupational groups. The hierarchy here runs from institutions which are training for the learned and scientific professions, through colleges tuned to the band of middle-class business and professional service occupations, to teacher training institutions and to junior colleges that are primarily devoted to vocational specialities or remedial work. Institutions that have been restricted to Negroes, who for one or another reason tend to score at the bottom of academic promise indices, tend to produce the lowest mean scores for entering freshmen (or, for that matter, for graduating seniors).

That these differences exist is pervasive knowledge. However, the magnitude of the differences is not so well known. In most states, there are institutions where a student in the top 10 percent of his class would fall in the bottom 10 percent of another institution within that state. In the College Board metric, institutions may be found whose entering freshmen average 275 or 725; in terms of the conventional I.Q., the range would be from the low 80's to the high 140's.

What impact these differences have for a more conscious commitment by given institutions to specific talent-manpower levels is uncertain. Thorndike and Hagen[76] have charted talent ranges (or, in our context, academic promise ranges) for a wide spectrum of occupations, but these data, as well as an earlier observation by D. E. Super[77] and the still earlier data on draftees in World War II,[78] show that each occupation has a relatively wide band of ability within it, and that the relationship between occupational field and mental ability does not so much stem from a relationship between ability and occupational success, but rather from a relationship between ability and sustained study beyond high school (together with the fact of educational requirements, meaningful or not, for some occupational fields). The picture is further complicated by findings such as those of James Davis,[79] who provided evidence, in his follow-up study of college graduates from a number of institutions, that the brightest students in each institution generally majored in science, but that taking the totality of institutions, science majors represented a wide range similar to the institutional differences themselves. At this point, and probably at future points, one cannot say with any assurance what minimum ability levels are needed for different manpower specialities. Accountants and doctors as groups have higher mental ability averages than file clerks, who do better on tests than lumberjacks; but some lumberjacks become doctors, and some doctors might make better lumberjacks. What can be said most precisely is that given the varying dependence among occupational fields on academic content, together with the traditional teaching methods involving symbols (verbal or numerical) and abstractions, the rather fuzzy hierarchies will probably continue. We need to expand the inventory of abilities and determine their precise occupational relevance before stating that a given training facility should admit students in given ability ranges. Yet, efforts with expanded aptitude or ability tests thus far have given no indication that this is possible. It is more likely that men will never discover nor agree on a precise constellation of attributes for any given field, but they must make some continuously revised decisions as to the style and content of training. As a social animal, man needs enough communality to communicate with others, particularly those with whom he works; but what is needed in a given field that one worker cannot provide, another colleague will. The point is this: there is no evidence from measurement studies that establishing training facilities to cater to certain talent ranges, and manipulation of the size of these facilities to control flows of trained manpower, is a reasonable possibility. Rather, it would seem that one should start from the other end. What content or skills (including academic or mental skills) are needed, and what alternate training procedures are likely to bring students, among whatever levels available, to mastery of that content? What diversities does an occupational field require? In this work, measurement is more likely to be useful as a tool to define proficiencies and to attest to when they have been reached than to answer completely the question of who shall be trained in what. The more precise proficiency criteria may lead to the development of more precise promise criteria. At the moment, we have a general promise criterion for a general academic proficiency criterion. The faster the learning pace, the more abstract it is, and the more memorizing of abstractions, facts, or images is required; the more reading or problem solving involved, the more complex the subject matter; and the more traditional the system of instruction, the greater the level of traditional mental ability that is required.

It has been stated that the general as well as the academic public perceives a diversity and a hierarchy (though dimly) of institutions in terms of levels of mental ability of entering students, and that this hierarchy does indeed exist. Another perception among the general public and most of academia is that the quality of an institution is synonymous with--or, at least, highly related to--the quality of entering freshmen. It is reasonable to assume that the higher the mental ability level of students, the faster is the pace that can be set, the richer is the content that can be required, and the greater the levels of competition that can be nourished--all of which imply a quality capability. But reason would tell us that after rather than before measures are needed (or better yet, both before-and-after measures) to determine what an institution has done for or to its students. Using admissions prerequisites as standards is a dean's way (instead of a faculty way) of controlling--or attempting to control--a faculty, who if they failed all students would have nothing much left to do. This logic suggests that our perception of the greater quality of Ivy League vs. struggling church-related college, for example, is more rationalizational than rational.

Added to the fact that it is possible to take good student raw material and do nothing much with it, thus casting doubt on this input characteristic as an infallible quality criterion, there is the differential attrition rate among institutions. From occasional descriptive reports or records, it is apparent that some institutions fail only 2 or 3 percent of their students, while others may drop 50 percent or more before the sophomore year. In other words, grading standards vary and are another traditional way of attempting to maintain and control the quality of the institution or program. Yet, no competent studies across institutions are known that shed much light on the forces behind differential attrition standards.

Generally speaking, records show that highly selective institutions at one extreme, and institutions hungry for students at whatever cost at the other, tend to fail relatively small proportions of enrolled students. Davis' data on students majoring in science[80] would indicate that attrition " standards" are a function of the institution rather than of absolute performance levels or field-related standards. Studies within single institutions experiencing rapid change in the level of entering students over a period of a few years (e.g., the studies by S. C. Webb[81] at Emory and Aiken[82] at the University of North Carolina) show that faculty members tend to maintain what they perceive as a going rate in assigning F's, whatever the fluctuations from year to year in ability. This and foregoing observations suggest that attrition rates are basic to the institution rather than to the student or major field, and that the need to retain students or a belief in the infinite superiority of the student are factors that depress attrition rates. It would seem that some important and intriguing studies could be conducted which might seek out faculty, administration, board of control, or constituency or manpower factors which affect the establishment of various attrition standards. Does overcrowding by an admissions office press faculty into thinning the ranks, or does it signal that more students are to be given the higher education treatment? What part is played by manpower needs? Has the acceptance of the Negro in a wide variety of high-level jobs put pressures on predominantly Negro institutions to graduate larger proportions of entering freshmen?

Crucial in cross-institutional studies of the sort that might deal with such questions is the further development of proficiency or academic attainment measures. Davis[83] has argued for a national grading system, although such would be quite threatening to many colleges (given the relationship between academic ability and performance, and the diversity of ability ranges over institutions), and it is difficult to imagine how such a system could be established. Another problem is the common variance--or overlap--between tests of ability and tests of achievement (for example, the National Teacher Examinations are more ability tests in disguise than tests covering content of teacher preparatory programs or of the skills useful in teaching).

It was noted earlier that personality tests grew principally out of the early vocational guidance interests and the mental hygiene movement of the 1940's. It was also noted that tests of personality have been used, but have yielded little, in studies of entering freshmen.

This refers, most exactly, to the use of personality measures in attempts at prediction of academic performance.

In a search for non-cognitive factors such as motivation, perseverance, stability, and the like which may affect academic performance, a virtual legion of investigators have tried each test as it comes out; there are more than a hundred studies reported, and probably many more unreported, that have examined the relationship of the Minnesota Multiphasic Personality Inventory to grades. Such work may be summarized by the statement that the normal and abnormal psychologists concerned with human development have their theories and criteria, and the academicians concerned with educational development have theirs, and never, it seems, would the twain meet. An early attempt to break this deadlock was made by Fricke at the University of Michigan, who sought responses to a somewhat wild array of items (e.g., preference for poodles or German police dogs), and then observed if particular responses had implications for later academic performance. His studies, reported in the resulting test manual,[84] have generally proved of more use in Fricke's hands than in the hands of others at other institutions, a not uncommon occurrence. Another relatively significant attempt was the construction of the Omnibus Personality Inventory[85] by a team at the (then) Center for the Study of Higher Education at Berkeley, where the concern was expressly for contriving a set of measures useful for exploring problems in higher educational development. This instrument still reeks heavily of the clinical heritage, for it was built by selecting items or scales from existing personality inventories that were felt to be relevant.

Potentially more valuable instruments were the Edwards Personal Preference Schedule[86] and the Student Activities Index.[87] Both of these instruments were drawn from the fifteen normal needs defined by Murray in 1938,[88] and seem to have fairly direct relevance to the efficiency of students' functioning in the academic situation. The Student Activities Index was a parallel instrument for another (the College Characteristics Index)[89] established specifically to study how the press of the college environment might serve the needs of students or frustrate them. Although these devices may have been carelessly or casually used in most of the studies conducted with them, not a great deal has come from either in defining important differences among entering student bodies.

Worthwhile results over a range of institutions have been reported by G. E. Schlesser[90] with his Personal Values Inventory, a personality measure designed specifically for predicting grades in college. However, not many of these reports are readily available in the literature, and the fragments one finds here or there imply that Schlesser's failure to include high school performance in his studies (arguing that the

PVI is a substitute for the uneven grading systems from school to school) may mean that the PVI is, in effect, a self-report of high school performance.

One of the more promising leads, with regard to measures of personality applied to the student input situation, has come from several investigators who have looked at the differential personality profiles of groups of students applying to various kinds of colleges. Several studies by the National Merit Scholarship Corporation's research team,[91] and some at Berkeley,[92] have shown that there are reasonable differences in such areas as achievement motivation when one looks at high-ability students who select competitive as opposed to non-competitive colleges. This may suggest that powerful college image factors perform a kind of sorting function, and that with the restriction of range on the more elusive personality factors their import for achievement or development is clouded in studies of single institutions. To complicate matters even further, it is reasonable to assume that a given personality trait--e.g., introversion--could facilitate achievement in one kind of program or institution and retard it in another.

The way out of these dilemmas would seem to involve, first, a more searching attempt to define personality constructs relevant to the academic demands and to learning (e.g., who has produced a good measure of interest in ideas?), and then to see these applied over a variety of institutions in conjunction with specific analyses of their image and their challenges. The differences in mood, if you will, in student bodies at Antioch or Ball State are apparent, and although the college atmosphere plays some small part it may be expected that image and input play a larger one in shaping it.

Although their instrumentation is not yet ready for general testing and application, it would be amiss not to cite the tack being taken by a measurement-oriented group of personality theorists, centering most exactly on Messick[93] at ETS. This group started with a concern for studying the error that seemed to systematically affect the performance of certain individuals on tests; for example, it was found that tests with items answerable in yes-no fashion were affected by a trait labeled "acquiescence"--acquiescent individuals tended as a matter of course to prefer a "yes" response, while the non-acquiescent tended to favor a "no." After some brief effort to establish test formats that would not be biased by some such response styles, Messick and his associates began to look at the styles as important human traits in their own right. What is emerging from this study is an inventory of problem-solving styles, related to the "cognitive styles" of other investigators, and some procedures for determining their relevance in a variety of situations. The general tenor of the developing theory is that different disciplines, or learning tasks, may involve or favor different problem-solving styles; understanding how these processes may function could afford talent identification and selection devices, but could better provide a basis for diagnosis and specific training in the appropriate styles needed. Although this may be the measurement scientist's way of legitimizing the popular belief that successful mental activity in mathematics is not the same as that in English Literature, there may be promise for some important breakthroughs in measurement and in the psychology of learning.

A new and relatively rich area of studies of student input have focused on relatively simple biographical kinds of factors--the student's socioeconomic class, the nature of his family and school experiences, and his attitudes on educationally relevant issues. The crucial modern work started with Clark and Trow,[94] who attempted to define student subcultures from biographical factors reflecting their purpose in going to college. These efforts led to the construction of the College Student Questionnaires[95] by Peterson, who has reported some of the differences observed over a range of institutions in their entering freshmen.[96] Within the past two or three years, Astin[97] and the American College Testing Program research staff[98] have also assembled some information on how the entering students vary in background, attitudes, and aspirations. Attempts to develop scales of such items reflecting psychological or developmental traits are still in very primitive form (e.g., Peterson has brief scales of such traits as family independence, cultural sophistication, and the like); but this may improve soon. The crucial aspect of the matter is that there has been strong interest among institutions in looking at their student bodies in such an intensive fashion. Knowing one's students better, particularly in these times of rapid change in college attendance, would seem a promising basis for better-directed educational treatment.

In regard, then, to an examination of measures of student input, the last decade has seen a more open recognition of the diversity in intellectual ability that marks American institutions of higher education. But more important, we are beginning to recognize that a host of other characteristics may have relevance for the educational or institutional mission, and we are beginning to be less concerned with seeking a preconceived quality of academic aptitude (defined unidimensionally) and more concerned with understanding some of the more subtle variations among students attracted to one or another kind of college. A formal examination of the attitudes of entering students may also provide a better basis, eventually, for assessing change in areas beyond the reach of hidebound achievement tests. But of greater importance, we may be just entering an era when training can be tailored to the student, rather than reserved for those who learn conventional things in conventional ways.

## The Enlarged Concern with the Educational Context

The measurement mainstream in higher education grew (some would say to flood stage) with a focus on the measurement of human traits and qualities through tests designed to tap some underlying continua. Whatever the specific goals of any institution of higher education, it must have as its major function the guiding and enhancement of individual growth and development. Measurement tools for such qualities as learning readiness, scholastic aptitude and achievement, learning styles, and the like have obvious utility for managing and monitoring some basic concerns of colleges and universities. Indeed, the means and dispersions of scores of student populations may express important institutional qualities.

However, having such measures and a background of social science research strategies, a person concerned with institutional functioning should now seek some ways of measuring institutional qualities with the aim of finding associations between what the institution is or does on the one hand and what happens to students on the other. Yet, until the last ten years, not much happened in the way of defining dimensions of institutional diversity based on institutional qualities, or in constructing scales to reflect these diversities.

Several factors may account for this surprising omission. Typological categories formed on such bases as control, type of curriculum, size, and the like have been used for some time, and serve identification and record purposes. There are popular tendencies, even among professionals, to view colleges and universities along a unitary continuum of goodness-badness (although simple reflection will indicate that an institution must have many different qualities of goodness and badness, or that institutional qualities good for some students may be bad for others). Most measurement specialists cut their professional teeth on the study of differences among individuals, of statistics and methodology, or both. Statistical procedures such as analysis of variance permit one to look at the possible influences of different situations on the student, thus reducing the necessity of having continua of institutional qualities. Persons concerned with the theory and philosophy of higher education have, on the other hand, seldom come from the disciplines of statistics and methodology or of measurement, and if they do they are frequently pressed in their professional roles to grapple with urgencies of housekeeping such as funding, arbitrating among various proponents of curricular change, and the like, leaving little time for such a staff function as measurement-based research.

For some time there have been two basic kinds of approaches to institutional analysis and assessment that are relevant to what we have called here the educational context. The first is essentially the case-study method, conducted by one or several perceptive observers with some variety of experience over a number of institutions. A good example of the results of this method is the set of ten studies provided by David Boroff in Campus, USA.[99] Measurement as such plays only the slightest of roles in this approach, and the consumer is at the mercy of the ability, wisdom, and experience of the observer.

Another time-honored approach to studies of the educational context in institutions of higher education is that typically employed by accrediting commissions. This may include case studies by the natives or visiting specialists, but it also hinges on the collection of descriptive data on characteristics assumed to be necessary components of a favorable learning environment. A library is judged to be an essential component; hence, the number of volumes, circulation rates, or the budget for continuing acquisition form indices that may be contrasted among an array of institutions, and used as a basis for qualitative judgment and standards.

Although this procedure grapples with some subtleties, it has three serious flaws. The first is that it must be based on the norms of current realities. Who is to say, in any absolute terms, the number of library volumes necessary for a given discipline or institution, other than someone with one eye on the going rates? With the advent of the computer as an instructional and research tool, should it not be incorporated into the minimum educational essentials? Can it be so incorporated until more institutions acquire one? What a college requires to fulfill its function should be the result of a careful series of studies to determine exactly what can be done with different resources. One remembers too well Anne Roe's[100] finding that the small midwestern college, typically without extensive laboratory equipment, is the major spawning ground for eminent physical scientists, or that Russian schools were able to accomplish a great deal with homemade laboratory apparatus.

A second serious flaw, not unrelated to the first, is that what an institution has, and what an institution does with what it has, may be two different things. There are those persistent critics who point out from time to time that this or that exemplary educational facility houses a complacent faculty and a thriving country club of students.

A third flaw frequently inherent in this approach has to do with the complexities of qualities and goals of institutions of higher education vs. the fact that what is generally observed comes from things readily observed or counted. We can, without much difficulty, count the number of Ph.D.'s on the faculty, or student contact hours; we can contrive ratios that seem also to have some value. But the point is that too frequently this approach has taken as base data the things that are available and that a registrar's clerk or president's administrative assistant can assemble. The lesson from measurement research is that one must _first_ grapple with a definition of essential qualities, _then_ seek ways of measuring them, and _then_ test their true meaning through systematic analysis against other criteria.

An important adaptation of this approach toward attempting to systemize procedures, defining group-related characteristics, and studying their meaning grew out of the work of a team of measurement specialists at the National Merit Scholarship Corporation. This effort, reported by Astin in 1962,[101] involved taking some thirty conventional kinds of indices for a group of more than three hundred colleges, and using factor-analytic techniques to determine the basic underlying dimensions of diversity therein. This study revealed, in Astin's interpretation, six principal dimensions: affluence (wealth), size, public vs. private control, masculinity vs. femininity, realistic or technical emphasis, and homogeneity. The largest proportion of the variances among these institutions had to do with the affluence factor, which was made up of such indices as measures of the college's financial resources, ability of students, proportion of Ph.D.'s on the faculty, and so on.

The procedure of factor analysis is limited, of course, by the fact that one can only find variances emerging from the elements studied that were already contained in the measures going into the analysis. But the components of ("loadings on") a factor reveal relationships among ingredients as well as provide a parsimonious or efficient way of dealing with a host of imperfectly related variables. J. M. Richards and others at the American College Testing Program have extended this treatment to the junior college (finding six factors or dimensions labeled cultural affluence, technological specialization, size, age, transfer emphasis, and business orientation).[102]

Although this kind of beginning is tremendously attractive to the measurement statistician at first, on second blush one must return to the possible dimensionality put into the system, and whether, as in the accrediting commission approach, one has chosen initial measures because they are important or simply because they are available. One might develop an efficient way of describing food by taking a number of measures of the contents of a grocery store and emerge with a packaging factor, a spoilage factor, a wet-dry factor; but the practical purposes of the shopper make the original distinctions of fish vs. fowl more important in provisioning for Friday's dinner. Blind searches may stumble on some useful leads, but a guided search may be more effective. In the latter case, it would seem we are less likely to end up with minor modifications of things we already knew.

The most significant advance in measuring institutional qualities (not in terms of what it has yet produced, but in terms of where it is now leading many investigators and theorists) can be attributed to C. R. Pace and G. G. Stern, who developed in the late fifties an instrument called the College Characteristics Index.[103] They reasoned that the important educational forces in the learning environment might best be revealed through the eyes of students, and they collected a number of statements concerning qualities or conditions that students might react to as generally true or not true for their campus. The content of these statements reflected perceptions of the competitive scholastic pressures, the status of the instructors, the topics of free or informal discussion among students, the emphasis on athletics, and the like. Application of the resulting instrument over a number of diverse campuses allowed the researchers to determine what items reflected prevalent differences, and how these differences might be grouped into scales.

The next major development of this approach grew out of an honest split which shortly developed between Pace and Stern. Stern, holding more to the interests of the personal and social psychologist, and heavily influenced by the need-press model of H. A. Murray,[104] turned the resulting scales on the College Characteristics Index toward measures that seemed to correspond with Murray's theoretical model and which were concerned with the total developmental needs of students and the parallel forces in the environment required for their satisfaction.[105] Pace, more concerned with institutional dimensions that might have useful meaning for the general administrator or faculty member, used a portion of the items, applied factor analysis to institutional means (rather than to individual student variance), and derived scales which he argues reflect true _institutional_ qualities. His instrument, called the College and University Environment Scales,[106] yields scores on five dimensions: (1) _practicality_: the degree to which personal status and practical benefit are emphasized in the college environment; (2) _community_: the degree to which the campus is friendly, cohesive, and group-oriented; (3) _awareness_: the degree of emphasis on self-understanding and personal identity, a wide range of appreciations, and personal involvement with the problems of the world; (4) _propriety_: the degree to which politeness, protocol, and consideration are emphasized; and (5) _scholarship_: the degree to which competitively high academic achievement is evidenced with concern for scholarship and intellectual discipline and interest in knowledge and ideas.

The stage would now seem to be set for studies that might attract many researchers to test for relationships between individual growth and institutional characteristics. Perhaps partly because this development is so recent and because early work in the development of instrumentation has to do, in essence, with tidying up the internal characteristics of items--or perhaps because the major initial use has been more by administrative observers than by measurement researchers--not a great deal has yet been reported on the validity, or tested meaning, of the scales. Some potential problems are apparent, however. One is that students are simply not aware of many important features of the college or university--for example, facilitation of faculty research. Another is that there may be some important omissions of content in the totality of items themselves, or that items get outdated rapidly, or that it is difficult to find items that provide as fair a set of stimuli for, say, a small, predominantly Negro college in the South as for an Ivy League university. Still another is the fact that the items are more oriented toward the perceptions of the student than toward his actual behavior (e.g., the distinction between "A lecture by an outstanding scientist would be poorly attended" and "I have attended a lecture by an outstanding visiting scientist during the past term").

A somewhat different approach from that of Pace and Stern was reported by Astin and Holland[107] in 1961. They reasoned (and not entirely without evidence) that those things which comprise the educationally relevant personal characteristics of student bodies, together with the students' relative emphases (through their majors) on different course areas, can be used to constitute the environment. Their procedure was to assemble information on average academic ability, institutional size, and the proportion of students majoring in departments grouped into six different areas--the latter kind of characteristic having been found to be quite stable for institutions over time. Tests of this approach against the College Characteristics Index provided some evidence that student perceptions of the environment are not unrelated to who the students are when they enter college as well as to where they place their major academic interests.

This procedure, called the Environmental Assessment Technique, could be employed by going to data of public record; that is, its components were made up of variances that although already known could perhaps be organized more efficiently for purposes of institutional definition. At least one major validation study[108] did explore the meanings of the scales against the students' perceptions of the effects that the individual colleges have upon them. From one perspective, the Astin and Holland studies give some useful insights about meanings of selectivity and programatic foci; from another, one might say that brighter students spend more time studying and are more likely to choose a humanities major than an education or business major, so ho-hum. Crucial for our purposes is the fact that some important new efforts were made; there was sufficient study of the results so that the limitations of findings could be recognized and the search could be pressed further.

In a newer series of studies done as part of a major new research program at the American Council of Education, Astin[109] has taken particular note of the perception vs. actual behavior question, and has developed some environmental assessment scales based on what students say they actually do. These are now being used in a systematic multi-college, multi-goal continuing study that most assuredly bears careful watching.

The most complex approach, and certainly the one with the most comprehensive attempt to build on some theoretical model of institutional functioning, is growing out of a series of activities led by Earl McGrath at the Institute of Higher Education at Teachers College, Columbia University, and involving a higher education research team at Educational Testing Service. Little more than an occasional working paper[110] has been released on this project as yet, but the effort is so potentially significant that it would be amiss not to attempt a summary of it here.

This effort, as defined by McGrath's team, has been greatly influenced by John Gardner's theory[111] of institutional functioning and self-renewal. Adopting a focus on "institutional vitality," they have made a variety of attempts to define its essential nature or natures and its components. These attempts have ranged from case studies of prototype institutions to the systematic polling of educational researchers or educational leaders. Although the project's main concern centers on forces in effective innovation and institutional continuance and survival, rather than on the more orthodox kinds of dimensions of previous environmental assessment work, the developing studies are beginning to provide some entry into the dynamic interrelationships among educational and administrative forces. One by-product that may shortly be available for wider use is an instrument that may be used to supplement student perceptions of the environment with faculty perceptions, and provide some additional dimensions that relate to academic freedom, receptivity to new ideas, etc.

The important thing about all of this work of the last decade would seem to be that serious attempts are being made to define and measure significant social, personal, and educational forces that may characterize institutions. The most important recognition of the decade may have been that of Pace, when, through his College and University Environment Scales, he said, "Let's measure the institution, not the students." There are those who may foresee a simple institutional evaluation function and expect a single good-bad dimension, thus questioning the practicability (because of threat) of applications; yet as is so often the case, careful attentiveness to an institution, as to an individual, shall surely reveal interesting and attractive complexities, and provide leads for understanding and effective modification in desirable directions. This more mature and acceptant attitude will surely prevail, carried on not only by the mushrooming concerns with educational technology and research, but also because it has its own intrinsic rewards.

## The Measurement of Impact of Colleges Upon Their Students

Not much will be said here on the measurement of the college's impact upon its students, partly because so little has been done that concerns direct measurement of outcomes, and partly because of an excellent and exhaustive review by Newcomb and Feldman[112] that is now in preliminary draft. That study, which I believe will be the most important milestone in higher education research since Sanford's 1962 contribution, covers work in progress as well as work reported, and provides a comprehensive frame of reference for future studies.

The problem in studying college impact is that it is extremely difficult to contrive a "clean" research design. One needs before-and-after measures, some ways of controlling personal (as opposed to context) factors, some way to separate the impact of the times from the impact of the institution, and some way to separate simple maturational effects from those produced by contrived educational intervention. Another problem is how impact will be defined and measured. Simple tests measuring achievement at the end of college are available, yet some normative studies of students year by year show a lowering (as introductory courses fade into the background) rather than a raising of scores with continued time in college; one begins to question not only the difficulty in contriving subject-matter tests fair to all students, teachers, and institutions, but also the validity of the academic way. There are those who call for the ultimate criterion of social, personal, and professional achievement in life itself, and who have found little or no relationship between academic success and success in life. The problem as to how the great mass of the higher educated, going into a variety of work and other societal roles, can be evaluated is a most difficult one that will probably be resolved, eventually, by the separate formulation of a variety of measures for a variety of purposes. Some of these measures may be generally relevant to the role of the educated adult (e.g., "social conscience"), and some may be peculiar to a particular subgroup of the population. Here again, even within occupational fields, we shall learn that there are few ultimate, universal qualities of goodness; we need a variety of kinds of doctors with a variety of kinds of skills or sensitivities.

One way around these dilemmas is to largely ignore the specification of a variety of impact criteria, and to note successful completion of an educational sequence and movement into a new educational or life sequence as the true test of development. Perhaps, indeed, after a half-century of testing and articulation studies, this is about all that we have accomplished. This is not a suggestion that we abandon the effort to specify goals; but it is a suggestion that we may do better to leave such efforts to the micro-view of the impact of the particular course, instructor or peer, or protest movement, rather than to the macro-view of what the total four years at a particular institution of higher education may have done to or for the individual.

Another way, which is essentially the strategy implied in the forthcoming volume by Newcomb and Feldman,[113] is to expand our knowledge of the various educational inputs, processes, and context, and, as Stein[114] has foreseen, to focus on the study of the interactions among environmental forces and student traits. Personal life or civilization itself is a series of progressive developments or adjustments of higher complexity to, hopefully, some higher order of functioning. It is my belief that at the very least we have reached a stage in higher education where the researcher, rather than needing to demonstrate that members of the faculty at a particular institution do not recognize, and variously disagree with, the goal statement in the catalogue, is and should be concerned with illuminating and specifying the many forces that make up an institution, and how these forces interact with one another.

## Directions for Future Research and Development

Measurement specialists could retire comfortably to the rose gardens behind their laboratories now if the implications of the history of the development of selection practices, and the supporting research, had given us an inventory of qualities important to subsume in measurement devices, and a clean avenue for their acceptance and use. Some people are hidden away working on such devices; some of the ideas for the devices have come from intuition, some from analysis of the literature of one or another domain within psychology, and some from exhaustive factor analyses of existing tests. But if we take the lessons of the past seriously, we must predict that little will come of the game of saying, "Here's a new concept that I know is important, so let's measure it and quit for today."

There are, however, some areas of considerable promise for the further development of the art and science of measurement, and for the educational system as a whole. The first of these has to do with defining, elaborating, and improving the criteria by which students are evaluated. For higher education, this has implications for the impact of the college on the student as well as for the transition of students from high school to college. This problem goes beyond noting what our present predictors in admissions studies tell us about the nature of the criteria, or the limitations of the criteria, although this may be a point of entry for the measurement specialist and a point of departure for the new studies. My own belief is that the problem calls for new approaches and new partners. We are now dealing not with the best scholar's definition of a subject-matter field, but with his intuition, as well as the intuitions of the significant societal leaders, as to the social and cultural utility of that field.

In our criterion construction, we must go beyond assembling a core of experts for a three-day conference, or beyond asking the faculty or societal leaders at large what the infinitely desirable qualities of man may be. The job probably starts with the best measurement specialists and the best teachers working together with a variety of students in an on-going instructional situation, which focuses on a product-by-product evaluation and re-evaluation. Where the content of the new criteria dimensions seems flimsy, the process then becomes that of confronting mind and spirit with the search for more substantial qualities. It is also suspected that in this work we shall not come out with a list of related qualities of goodness, but with a variety of frequently conflicting qualities. The latter, however, is probably the better model for meeting the various role demands of society.

A second area of promise, and one that is beginning to attract considerable attention, is that which Stein[115] called the "transactional" approach. He summarizes this by stating:

> Basic to this approach is the assumption that success in college, as all behavior, is a function of the transactions between the individual and his environment. Individuals affect and are affected by their environments. Consequently, for purposes of prediction it is important to understand both the characteristics of the individual and the environment.

We are experiencing a modicum of initial success with new environmental assessment techniques as well as with theoretical and operational studies of student characteristics other than intellectual ability. The work here is very much in the elementary stages of development, but it appears extremely promising. It not only subsumes convictions about selecting the student for the environment, but also for modifying the environment so that it may better serve the student. Some work with young children has shown that one can use tests to form subgroups of students who can be trained by different methods to the same criterion. If we can achieve this with older students and with some of the more advanced kinds of studies, then our educational institutions may support democracy in the more vital ways that democracy has always supported education. But whether we use such work to determine the kinds of environments in which a group of bright students learn best, the different kinds of environments in which different subgroups of bright students learn best, or the set of environments that may permit some subgroups of students lower on the traditional ability hierarchy to learn as well as brighter subgroups in conventional systems--any of these applications, could they be carried off, would seem worthwhile.

The appeal of the work on cognitive styles (or "problem-solving styles") is attractive and promising for these kinds of reasons. One can argue that different disciplines require different kinds of solution strategies, and that effective education is that which teaches the student new modes of attack. But here, again, the work must proceed hand-in-glove with the best subject-matter people the measurement specialist can muster to join him.

An important part of looking at promising avenues indicated by past research is to recognize that promise, defined by the measurement specialist, is not enough. Neither has educational practice always accepted the products of the measurement specialist only for the most noble or pure of reasons. What chance for useful contribution, indeed, does this leave him? Certainly with more than rushing to market a new test of creativity (one can be assured there would be buyers, particularly among the naive, and that the naive would be particularly vulnerable to unquestioned acceptance of the definition perpetrated by the test).

Measurement research is maturing to the point where now it may be more mission-oriented than discipline-oriented. The achievement, in our current academic world, of a mission orientation is being brought about by the multi-team approach. McGrath, had he lived two centuries ago, would have retired to a monkish cubicle and pontificated; today, his effort to define institutional vitality is involving literally thousands of people and hundreds of perspectives, each carefully chosen and managed. The statistical specialist, although he may want to hide and develop a few dozen new mathematical models, is being pressed into practical service by the idea man with a problem; the test constructor must sit down with his client on the firing line. The experience of working together is beginning to show some promise of an ultimate common language and the prospect of real communication.

Far too many pages ago, we noted that in the beginning was Harvard. Is her light still shining in the darkness? It may be, but it would seem that we are rapidly passing the stage, through measurement research activity, in which goodness is defined by an epitome institution (where the success of graduates may be assured by the success of their parents or by a universal stereotype of awe that may greet any product). It would seem that we are moving toward an attempt to learn by studies across all institutions as well as by focusing, where necessary, on discrete units within an institution, as McKeachie's paper[116] in this series, for one, demonstrates forcefully.

Measurement research in higher education may have contributed, most precisely, a relatively unitary dimension (scholastic aptitude) that has become an exclusive focus in admissions practices. This dimension was accepted because it saved time of faculty members, who otherwise would have devoted hours to test construction and grading; and it has misled us into selecting into our systems those who can be taught with the least effort, involvement, or difficulty. There is, to some readers of this report, some "proof" for such an interpretation. But others, perhaps of different biases, will see that what has been developed is a point of view that there is utility in attempts to specify some essential individual and institutional quality, and to test its meaning in some precise ways by studying its implications against other measures, and by looking, in interactional studies, for more than simple associations. The outcomes may produce some tests for educational consumers--but more than ever before, the pressures are for evaluating research not so much by its statistical niceties but by its (measured, of course !) impact on educational leaders and educational practice. Those who can accept a mission orientation honestly, who can learn to talk with and use those specialists from disciplines other than their own, who can use tools for their proper function--those people are opening their eyes to a magnificent dawning for the most exciting period of educational development yet.

## FOOTNOTES

1. The first part of this paper was presented essentially in its present form at a conference on Selection Practices held at Grasmere, England, in April, 1967, co-sponsored by the University of Lancaster and the Institute of Higher Education of Teachers College, Columbia University.

2. The major source for this section is a comprehensive historical account of admissions practices until 1900 by E. C. Broome entitled "A Historical and Critical Discussion of College Admissions Requirements." Columbia University Contributions to Philosophy, Psychology, and Education, vol. 11, Nos. 3-4. April, 1903.

3. Ibid., p. 204.

4. F. Bowles, "The Evolution of Admission Requirements." College Admissions: The Interaction of School and College, p. 24-36. New York, College Entrance Examination Board, 1956.

5. Ibid., p. 27.

6. Ibid., p. 25.

7. C. M. Fuess, The College Board--Its First Fifty Years. New York, Columbia University Press, 1950.

8. College Entrance Examination Board, Annual Report, 1965-66.

9. Fuess, op. cit.

10. H. Chauncey, unpublished address to Association of College Admissions Counselors, Highland Park, Ill. November 13, 1947.

11. W. S. Learned and B. D. Wood, The Student and His Knowledge. New York, The Carnegie Foundation for the Advancement of Teaching, 1938.

12. College Entrance Examination Board, Annual Report, 1945, p. 45.

13. Chauncey, *op. cit.*

14. Learned and Wood, *op. cit.*

15. F. Bowles, *Access to Higher Education: The International Study of University Admissions*. Vol. 1. New York, UNESCO and the International Association of Universities, 1963.

16. H. Dyer, Review of "Admissions--College and University," prepared for the forthcoming 4th ed. of *Encyclopedia of Educational Research*.

17. D. Harris, "The Relationship to College Grades of Some Factors Other Than Intelligence." *Archives of Psychology*, vol. 20, No. 131. 1931. "Factors Affecting College Grades: A Review of the Literature, 1930-1937." *Psychological Bulletin*, vol. 37, p. 125-66. 1940.

18. J. A. Fishman and Ann K. Pasanella, "College Admission Selection Studies." *Review of Educational Research*, p. 298-310. October, 1960.

19. J. A. Fishman, "Some Socio-Psychological Theory for Selecting and Guiding College Students." In N. Sanford, ed., *The American College: A Psychological and Social Interpretation of Higher Learning*, p. 666-89. New York, John Wiley, 1962.

20. M. I. Stein, *Personality Measures in Admissions*. Research Monograph No. 5. New York, College Entrance Examination Board, 1963.

21. D. E. Lavin, *The Prediction of Academic Performance: A Theoretical Analysis and Review of the Literature*. New York, Russell Sage Foundation, 1965.

22. L. V. Koos, *Private and Public Secondary Education*. Chicago, University of Chicago Press, 1931.

23. Junius A. Davis and N. Frederiksen, "Public and Private School Graduates in College." *Journal of Teacher Education*, vol. 6, p. 18-22. 1955.

24. Audrey M. Shuey, "Academic Success of Public and Private School Students in Randolph-Macon Woman's College: I. The Freshman Year." *Journal of Educational Research*, vol. 49, p. 481-92. 1956.

25. C. C. McArthur, "Personalities of Public and Private School Boys." *Harvard Educational Review*, vol. 24, p. 256-62. 1954. "Subculture and Personality During the College Years." *Journal of Educational Sociology*, vol. 33, p. 260-68. 1960.

26. J. L. Holland, "The Prediction of College Grades from the California Psychological Inventory and the Scholastic Aptitude Test." *Journal of Educational Psychology*, vol. 50, p. 135-42. 1959. "The Prediction of College Grades from Personality and Aptitude Variables." *Journal of Educational Psychology*, vol. 51, p. 245-54. 1960.

27. J. W. Getzels, "Non-IQ Intellectual and Other Factors in College Admission." In K. E. Anderson, ed., *The Coming Crisis in the Selection of Students for College Entrance*, p. 23-28. Washington, American Educational Research Association, 1960.

28. Junius A. Davis, "What College Teachers Value in Students." *College Board Review*, No. 56, p. 15-18. 1965.

29. Getzels, *op. cit.*, p. 28.

30. W. Coleman and E. E. Cureton, "Intelligence and Achievement: The 'Jangle Fallacy' Again." *Educational and Psychological Measurement*, vol. 14, p. 347-51. 1954.

31. E. B. Page, "The Imminence of Grading Essays by Computer." *Phi Delta Kappan*, vol. 47, p. 238-43. 1966.

32. S. Klein and R. Skager, *"Spontaneity Vs. Deliberateness" as a Dimension of Esthetic Judgment*. Research Bulletin 66-14. Princeton, N. J., Educational Testing Service, 1966.

33. Junius A. Davis, "Non-intellectual Factors in College Student Achievement." In *From High School to College: Readings for Counselors*, p. 72-81, esp. p. 73. New York, College Entrance Examination Board, 1965.

34. D. R. Saunders, *Moderator Variables in Prediction with Special Reference to Freshman Engineering Grades and the Strong VIB*. Research Bulletin 53-23. Princeton, N. J., Educational Testing Service, 1953.

35. N. Frederiksen and S. D. Melville, "Differential Predictability in the Use of Test Scores." *Educational and Psychological Measurement*, vol. 14, p. 647-56. 1954.

36. E. E. Ghiselli, "Differentiation of Individuals in Terms of Their Predictability." *Journal of Applied Psychology*, vol. 40, p. 374-77. 1956.

37. J. W. French, *Manual for the Experimental Comparative Prediction Batteries*. Princeton, N. J., Educational Testing Service, 1964.

38. G. E. Schlesser and J. A. Finger, "Non-intellective Predictors of Academic Success in School and College." *School Review*, vol. 73, p. 14-29. 1965.

39. S. Messick, "Personality Measurement and College Performance." In *Proceedings of the 1963 Invitational Conference on Testing Problems*, p. 110-29. Princeton, N. J., Educational Testing Service, 1963.

40. Fishman, *op. cit.*

41. Messick, *op. cit.*

42. Dyer, *op. cit.*

43. Bowles, *Access to Higher Education*, *op. cit.*

44. Dyer, *op. cit.*

45. See, for example, J. G. Darley, *Promise and Performance: A Study of Ability and Achievement in American Higher Education*. Berkeley, Center for the Study of Higher Education, University of California, 1962. A. W. Astin, "Distribution of Students Among Higher Educational Institutions." *Journal of Educational Psychology*, vol. 55, p. 276-87. 1964. Astin, *Who Goes Where to College?* Chicago, Science Research Associates, 1965. R. C. Nichols, "College Preferences of Eleventh Grade Students." *NMSC Research Reports*, vol. 2, No. 9. 1966. A. W. Astin, R. J. Panos, and J. A. Creager, "National Norms for Entering College Freshmen--Fall, 1966. *ACE Research Reports*, vol. 2, No. 1. 1967. College Entrance Examination Board, *Manual of Freshman Class Profiles, 1967-69*. New York, CEEB, 1967.

46. Junius A. Davis, "The Criterion Problem in College Admissions Research." In J. M. Duggan, ed., *Research in Higher Education: Guide to Institutional Decisions*, p. 25-34. New York, College Entrance Examination Board, 1965.

47. K. E. Eble, *The Profane Comedy*. New York, Macmillan, 1962.

48. G. H. Hanford, "Testing: Wise Restraints." In L. Wilson, ed., Emerging Patterns in American Higher Education, p. 225-27. Washington, American Council on Education, 1965.

49. Davis, "Non-intellectual Factors in College Student Achievement," op. cit., p. 79.

50. J. W. Trent, "A New Look at Recruitment Policies." College Board Review, No. 58, p. 7-11. Winter 1965-66.

51. Learned and Wood, op. cit.

52. Junius A. Davis, Distribution of 1957 Entering Freshmen on Pre-Admissions Indices. Atlanta, Board of Regents of the University System of Georgia, 1958.

53. For a summary of this work, see D. G. Paterson and J. G. Darley, Men, Women, and Jobs. Minneapolis, University of Minnesota Press, 1936.

54. T. M. Newcomb, Personality and Social Change. New York, Dryden, 1943.

55. C. R. Pace, They Went to College. Minneapolis, University of Minnesota Press, 1941.

56. D. Chamberlin, E. Chamberlin, N. E. Drought, and W. E. Scott, Did They Succeed in College? New York, Harper, 1942.

57. P. L. Dressel and L. B. Mayhew, General Education: Explorations in Evaluation. Washington, American Council on Education, 1954.

58. B. S. Bloom, ed., A Taxonomy of Educational Objectives: Part I, the Cognitive Domain. New York, David McKay, 1956.

59. For a general description of this instrument, see S. R. Hathaway, "The Minnesota Multiphasic Personality Inventory." In O. J. Kaplan, ed., Encyclopedia of Vocational Guidance. New York, Philosophical Library, 1948.

60. J. P. Guilford and W. S. Zimmerman, Manual for the Guilford-Zimmerman Temperament Survey. Beverly Hills, Calif., Sheridan Supply Company, 1949.

61. P. E. Jacob, *Changing Values in College: An Exploratory Study of the Impact of College Teaching*. New York, Harper, 1957.

62. Sanford, *op. cit.*

63. A. W. Astin, *The College Environment*. Washington, American Council on Education, 1968.

64. T. M. Newcomb and K. A. Feldman, *The Impacts of Colleges upon Their Students*. Forthcoming.

65. M. Keeton *et al.*, a forthcoming volume probably to carry the title of the study, *The Future of Liberal Arts Colleges*.

66. B. R. Clark and M. Trow, "The Organizational Context." In T. M. Newcomb and E. K. Wilson, eds., *College Peer Groups: Problems and Prospects for Research*, p. 17-70. Chicago, Aldine, 1966.

67. Sanford, *op. cit.*

68. H. S. Dyer, "Can Institutional Research Lead to a Science of Institutions?" *Educational Record*, vol. 47, No. 4, p. 452-66. Fall, 1966.

69. W. H. Cowley, "Two and a Half Centuries of Institutional Research." In R. G. Axt and H. T. Sprague, eds., *College Self Study*, p. 17-22. Boulder, Colo., Western Interstate Commission for Higher Education, 1960.

70. Harris, *op. cit.*

71. Fishman and Pasanella, *op. cit.*

72. J. R. Hills, "Admissions Procedures that Make Sense." In Duggan, *op. cit.*, p. 16-24.

73. Learned and Wood, *op. cit.*

74. Darley, *op. cit.*

75. Astin, *Who Goes Where to College*, *op. cit.*

76. R. L. Thorndike and Elizabeth Hagen, *10,000 Careers*. New York, John Wiley, 1959.

77. D. E. Super, "A Theory of Vocational Development." American Psychologist, vol. 8, p. 185-90. May, 1953.

78. Naomi Stewart, "AGCT Scores of Army Personnel Grouped by Occupation." Occupations, vol. 26, p. 5-41. 1947.

79. James A. Davis, "Reference Group Processes and the Choice of Careers in Science." Paper presented at 1964 convention of the American Psychological Association in Los Angeles.

80. Ibid.

81. S. C. Webb, "Measured Changes in College Grading Standards." College Board Review, No. 39, p. 27-30. 1959.

82. Louis R. Aiken, Jr., "The Grading Behavior of a College Faculty." Educational and Psychological Measurement, vol. 23, No. 2, p. 319-22. 1963.

83. Davis, "Reference Group Processes and the Choice of Careers in Science." Op. cit.

84. B. G. Fricke, Opinion, Attitude and Interest Survey Handbook. Ann Arbor, University of Michigan, Evaluation and Examinations Division, 1963.

85. Paul A. Heist et al., Omnibus Personality Inventory Research Manual. Berkeley, Center for the Study of Higher Education, University of California, 1962.

86. Manual for the Edwards Personal Preference Schedule. New York, The Psychological Corporation, 1959.

87. G. G. Stern, Scoring Instructions and College Norms for the Activities Index and the College Characteristics Index. Syracuse, N. Y., Syracuse University, Psychological Research Center, 1963.

88. H. A. Murray, Explorations in Personality. New York, Oxford University Press, 1938.

89. Stern, op. cit.

90. Schlesser and Finger, op. cit.

91. For a summary of these studies, see "Tenth Annual Review of Research," NMSC Research Reports (vol. 2, No. 11, 1966), published by the National Merit Scholarship Corporation.

92. E. D. Farwell et al. "Student Personality Characteristics Associated with Groups of Colleges and Fields of Study." College and University, vol. 37, p. 229-41. 1962. P. A. Heist et al. "Personality and Scholarship." In W. W. Charters, Jr., and N. L. Gage, eds., Readings in the Social Psychology of Education, p. 65-73. Boston, Allyn & Bacon, 1963.

93. F. Dameron and S. Messick, Response Styles and Personality Variables: A Theoretical Integration of Multivariate Research. Research Bulletin 65-10. Princeton, N. J., Educational Testing Service, 1965.

94. Clark and Trow, op. cit.

95. R. E. Peterson, Technical Manual, College Student Questionnaires. Princeton, N. J., Educational Testing Service, 1965.

96. R. E. Peterson, Some Biographical and Attitudinal Characteristics of Entering College Freshmen. Research Bulletin 64-63. Princeton, N. J., Educational Testing Service, 1964.

97. R. J. Panos, A. W. Astin, and J. A. Creager, "National Norms for Entering College Freshmen--Fall, 1967." ACE Research Reports, vol. 2, No. 7. 1967.

98. College Student Profiles. Iowa City, Iowa, American College Testing Program, 1966.

99. D. Boroff, Campus, USA. New York, Harper Bros., 1961.

100. Anne Roe, "A Psychological Study of Eminent Physical Scientists." Genetic Psychology Monographs, vol. 43, p. 121-239. 1951.

101. A. W. Astin, "An Empirical Characterization of Higher Educational Institutions." Journal of Educational Psychology, vol. 53, p. 224-35. 1962.

102. J. M. Richards, Lorraine Rand, and L. P. Rand, A Description of Junior Colleges. ACT Research Report No. 5. Iowa City, Iowa, American College Testing Program, 1965.

103. C. R. Pace and G. G. Stern, "An Approach to the Measurement of Psychological Characteristics of College Environments." *Journal of Educational Psychology*, vol. 49, p. 269-77. 1958.

104. Murray, *op. cit.*

105. G. G. Stern, "Environments for Learning." In Sanford, *op. cit.*, p. 690-730.

106. C. R. Pace, *College and University Environmental Scales*. Princeton, N. J., Educational Testing Service, 1963.

107. A. W. Astin and J. L. Holland, "The Environmental Assessment Technique: A Way to Measure College Environments." *Journal of Educational Psychology*, vol. 52, p. 306-16. 1961.

108. A. W. Astin, "Further Validation of the Environmental Assessment Technique." *Journal of Educational Psychology*, vol. 54, p. 217-26. 1963.

109. Astin, *The College Environment*, *op. cit.*

110. R. E. Peterson and D. E. Loye, eds., *Conversations Toward a Definition of Institutional Vitality*. Princeton, N. J., Educational Testing Service, 1967.

111. J. W. Gardner, *Self-Renewal: The Individual and the Innovative Society*. New York, Harper & Row, 1964.

112. Newcomb and Feldman, *op. cit.*

113. *Ibid.*

114. Stein, *op. cit.*

115. *Ibid.*

116. W. J. McKeachie, "New Developments in Teaching." In E.H. Hopkins, ed., *New Dimensions in Higher Education*. No. 16. Washington, U.S. Office of Education, 1967.

## ANNOTATED BIBLIOGRAPHY

### A. The History, Philosophy, and Practice of Selective Admissions

1. Bowles, F., Access to Higher Education: The International Study of University Admissions. UNESCO and the International Association of Universities, vol. I. New York, 1963.

Drawing on his fifteen years' experience as president of the College Board, as well as upon a two-year study of university admissions in a number of countries throughout the world, this scholarly study relates education and its functioning to the public purpose, and treats the college admissions process as a broad, social phenomenon, responsive to deep-seated national pressures and aspirations within a country.

2. ________________, The Refounding of the College Board, 1948-1963. New York, College Entrance Examination Board, 1967.

"An informal commentary and selected papers" documenting the activities of the College Entrance Examination Board under Bowles's period of directorship. This period is particularly crucial, for its beginning marked the establishment of Educational Testing Service as the technical arm of the College Board. The collection chronicles the development of modern perceptions of selection needs and policies in the period of rapid development and application of admissions testing.

3. Broome, E. C., A Historical and Critical Discussion of College Admissions Requirements. Columbia University Contributions to Philosophy, Psychology, and Education, vol. XI, Nos. 3-4. April, 1903. (Reprinted 1963, College Entrance Examination Board.)

A classic and scholarly (in the old sense) study of the evolution of admissions practices as a function of the evolution of institutions of secondary and higher education. A "must" for the educational historian as well as for the college admissions officer.

4. College Entrance Examination Board, *College Admissions*. Vols. 1-10. One volume per year from 1954 to 1963; published in New York by the College Entrance Examination Board.

For a ten-year period the College Entrance Examination Board held an annual invitational conference for a small group of admissions officials on selected problems in college admissions. Speakers were chosen with special care; this set of volumes presents their papers.

5. Duggan, J. M., and P. H. Hazlett, *Predicting College Grades*. New York, College Entrance Examination Board, 1961.

A workbook for the non-initiated that provides cookbook formulas and worksheets for handling a prediction problem.

6. Dyer, H. S., "Admissions--College and University." To appear in R. Ebel and Victor Noll, eds., *Encyclopedia of Educational Research*. (4th Ed.) Forthcoming from the American Educational Research Association.

A perceptive review of problems, practices, and research in admission of students to college, this sweeping review is as appropriate for the layman as for the sophisticated researcher.

7. Fuess, C. M., *The College Board--Its First Fifty Years*. New York, Columbia University Press, 1950.

A folksy account by a former private school headmaster active for many years with the College Board, this volume describes (generally accurately, always delightfully) the first fifty years of development of the College Entrance Examination Board.

8. Thrasher, B. A., *College Admissions and the Public Interest*. New York, College Entrance Examination Board, 1966.

This book contains a series of literate reflections on the admissions process by B. Alden Thrasher, for twenty-five years director of admissions at M.I.T. Thrasher is particularly concerned with the social forces that push students into higher education. Theoretical and philosophical analyses are quite keen; the volume is less satisfying in its discussion of how these issues may be handled in practice.

## B. Reviews of Research on the Prediction of Success in College

9. Davis, Junius A., "Non-intellectual Factors in College Student Achievement." In From High School to College: Readings for Counselors, p. 72-81. New York, College Entrance Examination Board, 1965.

Although directed to pre-college counselors in the secondary schools, this general review concludes that there is not much beyond conventional aptitude and academic achievement for the high school counselor to use.

10. Fishman, J. A., and Ann K. Pasanella, "College Admission Selection Studies." Review of Educational Research, p. 298-310. October, 1960.

This paper, directed toward the moderately technical reader, reviews 580 studies during the decade from 1949 to 1959. Its useful bibliography contains fifty-seven references.

11. Garrett, H. F., "A Review and Interpretation of Investigations of Factors Related to Scholastic Success in Colleges of Arts and Science and Teachers Colleges." Journal of Experimental Education, vol. 18, p. 91-138. 1949.

This report is a review of prediction studies over a twenty-year period beginning in 1930, for which some 194 studies are mentioned.

12. Harris, D., "Factors Affecting College Grades: A Review of the Literature, 1930-1937." Psychological Bulletin, vol. 37, p. 125-66. 1940.

13. ________________, "The Relationship to College Grades of Some Factors Other Than Intelligence." Archives of Psychology, vol. 30, No. 131. 1931.

Both articles by Harris are reviews of prediction studies published over the period indicated. He focuses on the attempts to find correlates of achievement in college beyond the conventional cognitive measures. His summary indicates that those who do this kind of study are generally a rather haphazard lot.

14. Lannholm, G. V., "Review of Studies Employing GRE Scores in Predicting Success in Graduate Study, 1952-1967." Graduate Record Examinations Special Report, No. 68-1. Princeton, N. J., Educational Testing Service, 1968.

A review of some thirty-six published and unpublished studies which employed GRE scores in predicting success in graduate study.

15. Lavin, D. E., The Prediction of Academic Performance: A Theoretical Analysis and Review of the Literature. New York, Russell Sage Foundation, 1965.

The most recent general review of the literature on the prediction of student performance, this volume is noteworthy for its grasp of the broader issues as well as for its treatment of sociological and social-psychological factors that affect levels of achievement.

16. Stein, M. I., Personality Measures in Admission. New York, College Entrance Examination Board, 1963.

This report is an excellent summary of a review commissioned by the College Board of the use of personality measures in college admissions. Its major contribution lies in its analysis of current failures and in implications for future studies. This volume contains a useful bibliography.

## C. The Evaluation, Through Measurement Perspectives, of Higher Education

17. Brumbaugh, A. J., Research Designed to Improve Institutions of Higher Learning. Washington, American Council on Education, 1960.

Although measurement is not a matter of particular concern in this little guide, a useful handbook for the general administrator concerned with institutional self-studies is provided.

18. Dressel, P. L., et al., Evaluation in Higher Education. Boston, Houghton Mifflin, 1961.

This volume is a collection of essays by Dressel and his associates at Michigan State University on applications of measurement and evaluation procedures and points of view to the administration of the institution and its programs. It provides an excellent overview, as for a beginning graduate student in a formal study of higher education, of problems ranging from selection and placement of students, through evaluation of growth of students in various areas, to evaluation of instruction or institutional self-study.

19. Dressel, P. L., and L. B. Mayhew, General Education: Explorations in Evaluation. Washington, American Council on Education, 1954.

This volume summarizes for the intelligent layman the results of a four-year cooperative study of the impact of the general college programs in nineteen colleges and universities. Concluding chapters on implications and unresolved issues, and suggestions for future studies, are particularly noteworthy.

20. Lazarsfeld, P. F., and S. D. Sieber, Organizing Educational Research. Englewood Cliffs, N. J., Prentice-Hall, 1964.

Although drawn from the perspective of sociology rather than measurement *per se*, and frequently critical of measurement scientists (especially those outside the colleges), this volume is a useful and thoughtful summary of the problems in modern educational research.

## D. Current Standard Texts in Tests and Measurement

21. Chauncey, Henry, and John E. Dobbin, Testing: Its Place in Education Today. New York, Harper & Row, 1963.

Directed toward parents and teachers, this informative little volume cuts through many of the sources of popular fallacies and confusion about the place of testing in education today. A central notion has to do with the test as a "partner of teaching." Although directed toward the public school setting, many sections are quite useful for students of higher education.

22. Cronbach, Lee J., Essentials of Psychological Testing. 2nd ed. New York, Harper Brothers, 1960.

This book is far and away the most popular undergraduate text in general psychological testing.

23. Jackson, Douglas N., and Samuel Messick, eds., Problems in Assessment. New York, McGraw-Hill, 1967.

A monumental collection, through seventy-four chapters and almost a thousand pages, of classic studies as well as modern statements of contemporary issues. It is directed most precisely at the graduate student or measurement research specialist.

24. Lindquist, E. F., ed., Educational Measurement. Washington, American Council on Education, 1951.

Although now out-of-date (a revised edition is being prepared under the editorship of R. L. Thorndike), this volume is the classic reference for the planning, construction, use, and analysis of the educational test.

25. Linn, R. L., J. A. Davis, and Patricia Cross, A Guide to Research Design. Princeton, N. J., Educational Testing Service, 1965.

Focusing principally on tests available from ETS for general institutional research purposes, this manual is written for the institutional researcher who has entered that role from some background other than social science research or statistics. It includes appendices on statistical terms and procedures, and provides more than one hundred references.

26. Nunnally, Jum C., Jr., Tests and Measurements: Assessment and Prediction. New York, McGraw-Hill, 1959.

This volume, like the Cronbach volume, is a popular undergraduate text in test and measurement. It is particularly useful in terms of its concise and lucid treatment of statistical problems in the use of testing.

27. Stuit, D. B., G. C. Helmstadter, and N. Frederiksen, Survey of College Evaluation Methods and Needs: A Report to the Carnegie Corporation. Princeton, N. J., Educational Testing Service, 1956.

This out-of-print report is a summary of methods and materials for evaluating, in self-studies, the following aspects of a college: institutional objectives, curriculum, faculty, instructional effectiveness, student body, and student personnel services. After attempts to define the evaluation problem in terms of underlying dimensions in each area, the report cites both available and needed methods and materials.

## E. Surveys of Student Input Dimensions of Diversity Among Institutions of Higher Education

28. Astin, A. W., R. J. Panos, and J. A. Creager, "National Norms for Entering College Freshmen--Fall, 1966." ACE Research Reports, vol. 2, No. 1. 1967. Washington, American Council on Education, 1967.

Drawing from Astin's new concern with student behavior as an indication of the learning climate of an institution, this reference source provides a good answer to the question, "What are students like today?" The data come from a survey of almost 300,000 entering freshmen students at a carefully selected sample of 359 colleges and universities in 1967. Distinctions among various types of institutions are presented.

29. College Entrance Examination Board, Manual of Freshman Class Profiles, 1967-69. New York, College Entrance Examination Board, 1967.

Intended as a source book for secondary school counselors and others who help students make their college plans, the statistics on tested ability and high school performance of entering freshmen classes have served a number of research studies well. This sixth edition contains profiles supplied by 520 member colleges of the College Board.

30. College Entrance Examination Board, Manual of Freshman Class Profiles for Indiana Colleges. New York, College Entrance Examination Board, 1965.

This volume is significant because it attempts to demonstrate the feasibility of augmenting the College Board's National Manual of Freshman Class Profiles by including, in addition to test data and a self-description for each college, the results of formal environmental assessment studies and data on the kinds of students entering the college. It is published as a guide for high school counselors and students.

31. Darley, J. G., Promise and Performance: A Study of Ability and Achievement in American Higher Education. Berkeley, Center for the Study of Higher Education, University of California, 1962.

Darley's inquiry is concerned with the present structure of higher education in the United States in terms of ability, performance, attrition, and occupational plans of students in a national sample of institutions. Contrasts between freshmen in 1952 and 1959 are presented, with intensive analyses of students in Minnesota, Wisconsin, Ohio, and Texas.

32. Hills, J. R., Counselor's Guide to Georgia Colleges. Atlanta, Office of Testing and Guidance, Board of Regents, University System of Georgia, 1965.

An updating of an earlier survey by J. A. Davis, this manual presents admissions data on entering college freshmen, together with procedures for prediction of grades, in the public and private colleges of Georgia.

33. Learned, W. S., and B. D. Wood, The Student and His Knowledge: A Report to the Carnegie Foundation on the Results of the High School and College Examinations of 1928, 1930, and 1932. New York, Carnegie Fund for the Advancement of Education, 1938.

The classic study of educational development which examined achievement levels over secondary schools and colleges in the state of Pennsylvania by the use of educational tests, this investigation provided the first positive indication of the extent of the diversity that exists both among institutions and among departments within institutions.

34. Peterson, R. E., Technical Manual, College Student Questionnaires. Princeton, N. J., Educational Testing Service, 1965.

Intended as a guide for users of the College Student Questionnaires, this manual provides normative information on the range of background factors, aspirations, and experiences of college students.

35. Seibel, Dean W., A Study of the Academic Ability and Performance of Junior College Students. Princeton, N. J., EAS Field Studies Report, Educational Testing Service, 1965.

This study is a follow-up of a representative national sample of high school seniors for whom ability measures (from the Preliminary Scholastic Aptitude Test) were available. Data are presented which describe the academic ability of students who enroll in two-year institutions and students who enroll in four-year institutions according to their performance during the first year of college.

## F. The Analysis of the Learning Context or the College Environment

36. American College Testing Program, College Student Profiles. Iowa City, Iowa, American College Testing Program, 1966.

This volume, prepared by the ACT Research and Development Division, is an extensive description of students enrolled in colleges and universities using the ACT program. Statistical data include information on a wide range of student characteristics. A testimony to the diversity that exists among American institutions of higher education, this volume is of interest to anyone with patience to examine its data, and who is interested in contrasting his institution with others or is concerned with the broad range.

37. Astin, A. W., *The College Environment.* Washington, American Council on Education, 1968.

This report, heavily based on data collected through the National Merit Scholarship Corporation, is an excellent analysis of the problems and prospects in environmental assessment. Its implications are directed both toward the general college administrator and toward the individual psychologist or teacher concerned with the impact of the environment on human development.

38. ______________, "Distribution of Students Among Higher Educational Institutions." *Journal of Educational Psychology*, vol. 55, p. 276-87. 1964.

39. ______________, *Who Goes Where to College?* Chicago, Science Research Associates, 1965.

Astin has provided a useful guide for pre-college counselors as well as a set of environmental measures that are attracting wide use in cross-institutional studies. Using dimensions developed by the National Merit Scholarship Corporation team, Astin provides profiles for more than a thousand institutions. His dimensions are drawn from statistical combinations of facts of record about the institutions.

40. Barton, A. H., *Organizational Measurement and its Bearing on the Study of College Environments*. New York, College Entrance Examination Board, 1961.

This review is a sweeping and landmark summary, commissioned by the College Board, of sociologists' experience in environmental assessment, with implications for carrying the problems and procedures to assessment of the college environment.

41. Pace, C. Robert, *Analyses of a National Sample of College Environments*. Washington, U.S. Office of Education, 1967.

This volume describes work done by Pace in a USOE study of 175 colleges and universities in 1964 and 1965. Some useful contrasts between the author's approach and that of Astin are included.

42. ________________, College and University Environment Scales: Technical Manual. Princeton, N. J., Educational Testing Service, 1963.

This manual, prepared as a guide for users of Pace's instrument, contains the rationale and much descriptive data on the American college environment. A revised edition, now in manuscript form, will shortly be available; that edition will augment the original data by incorporating information from many of the institutions studied in the NORC follow-up of college graduates.

43. Peterson, Richard E., The Scope of Organized Student Protests in 1964-1965. Princeton, N. J., Educational Testing Service, 1966.

This volume reports the results of a study of institutional factors which are popularly assumed to have implications for incidents of student protest. A number of myths are exploded and a number of hidden factors are revealed. The study involved the majority of four-year colleges and universities across the country.

44. Stern, G. G., "Characteristics of the Intellectual Climate in College Environments." Harvard Educational Review, vol. 3, p. 5-41. 1963.

This paper probably represents the best review available today of the author's approach in measuring the college environment. Although semi-technical, it is presented in a form that the intelligent layman may understand.

## G. The Impacts of Colleges Upon their Students

45. Davis, James A., Great Aspirations. Chicago, Aldine, 1964.

46. ________________, Undergraduate Career Decisions: Correlates of Occupational Choice. Chicago, Aldine, 1965.

These two volumes are the first and second reports of a sweeping follow-up of graduates from some three hundred American colleges and universities. The data are of interest not only in their own right, but also because other investigators may obtain them from NORC after that organization has completed its initial analyses. As a data bank for a continuing inquiry, this is a highly significant effort.

47. Newcomb, T. M., Personality and Social Change. New York, Dryden, 1943.

This classic study found that young ladies from conservative Republican backgrounds moved in their views toward those of their faculty in a college where most instructors were liberal Democrats. It is significant for the solid evidence it presents on the more subtle issue of attitude change of college students as a function of their experiences in college.

48. ________________, and K. A. Feldman, The Impacts of Colleges Upon Their Students. Forthcoming.

This volume, now in pre-publication draft, is the result of a Carnegie-supported exhaustive review of a variety of research projects, both published and ongoing, that have implications for defining, measuring, and manipulating the impact of colleges upon their students. It may well become, when published, the most significant contribution to higher education research of this decade.

49. Pace, C. R., They Went to College. Minneapolis, University of Minnesota Press, 1941.

A follow-up of former University of Minnesota General College students, this study was an early classic toward providing information back to teachers and administrators that might be used to modify the higher education experience.

## H. Research Reports from Organizations Concerned with Measurement Studies of Higher Education

50. American College Testing Program, ACT Research Reports.

This series was begun in 1965 by staff of the Research and Development Division of the American College Testing Program in Iowa City, Iowa. Reports are issued as studies are completed. Generally they involve analyses of data collected in the ongoing programs of that organization toward particular problems, such as educational goals of entering freshmen, or the relationship between academic and non-academic accomplishment.

51. American Council on Education, ACE Research Reports.

Published from time to time by the Office of Research, American Council on Education in Washington, these reports draw principally on data from a major, ongoing, and multi-purpose study of a sample of colleges and universities.

52. College Entrance Examination Board, College Board Review.

This "slick" journal is published quarterly; its principal targets are college admissions officers and high school guidance counselors. Papers are literate and non-technical; important CEEB research projects as well as think pieces about the admissions or guidance process are included.

53. Educational Testing Service, Research Bulletins.

Viewed internally as a pre-publication issue, these bulletins report the results of major studies done by the ETS research staff as those studies are completed. The RB series is generally directed at the sophisticated researcher within the area of the subject of the study. A listing of research bulletins of general interest is contained in the ETS Annual Report.

54. National Merit Scholarship Corporation, NMSC Research Reports.

This excellent series was initiated in 1965 by R. C. Nichols. With a prospective phasing out of research activities at NMSC it is now threatened with extinction. Reports are directed toward the professional audience. See especially NMSC Research Report, 1966, vol. 2, No. 11, "Tenth Annual Review of Research," which contains not only work in progress and completed that year, but also provides abstracts of studies completed at NMSC from its founding in 1955. In all, 130 papers or projects are listed.

## I. Major Professional Journals Containing Reports of Measurement Studies

American Educational Research Journal

American Journal of Sociology

American Sociological Review

British Journal of Educational Psychology

College and University

Contemporary Psychology

EAS Resources (ETS)

*Education Recaps* (ETS)

*Educational and Psychological Measurement*

*Educational Record*

*Harvard Educational Review*

*Journal of Applied Psychology*

*Journal of College Student Personnel*

*Journal of Counseling Psychology*

*Journal of Educational Measurement*

*Journal of Educational Psychology*

*Journal of Educational Sociology*

*Journal of Psychological Studies*

*Personnel and Guidance Journal*

*Psychological Abstracts*

*Psychological Bulletin*

*Psychological Monographs*

*Public Opinion Quarterly*

*Register of Research Projects in Higher Education*
(Society for Research into Higher Education, Ltd., 2 Woburn Square, London, W.C.1)

*The Research Reporter*
(Center for Research and Development in Higher Education, University of California at Berkeley)

*Review of Educational Research*

*School and Society*

*Science*

*Sociology of Education*

## J. Organizations or Research Centers with Measurement Research Teams in Higher Education

American College Testing Program
(Iowa City, Iowa)

American Council on Education
(Washington, D. C.)

Bureau of Applied Social Research
(Columbia University)

Center for Research and Development in Higher Education
(University of California at Berkeley)

Center for Research on Learning and Teaching
(University of Michigan)

Center for the Study of Evaluation of Instructional Programs
(University of California at Los Angeles)

Center for the Study of Higher Education
(University of Michigan)

Centre for the Study of Higher Education
(University of Lancaster, England)

College Research Center
(Vassar College)

Educational Testing Service
(Princeton, N. J.)

Institute of Education
(University of London, England)

Institute of Higher Education
(Teachers College, Columbia University)

Institute for Social Research
(University of Michigan)

National Merit Scholarship Corporation
(Evanston, Ill.)

National Opinion Research Center
(University of Chicago)

# REACTIONS

In order for this second series of "New Dimensions in Higher Education" to better serve the needs of colleges and universities throughout the nation, reader reaction is herewith being sought. In this instance, with respect to Applications of the Science of Measurement to Higher Education, the following questions are asked:

1. Can you suggest other completed research, the results of which would add significantly to this report?

2. What problems related to this subject should be given the highest priority, in terms of further research?

3. What can the United States Office of Education do to encourage and support constructive innovation and change, based upon recent developments in the science of measurement?

Kindly address reactions to:

Dr. Winslow R. Hatch
Bureau of Higher Education Research
Office of Education
U. S. Department of Health, Education, and Welfare
Washington, D. C. 20202